

score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

OM nucleic - nucleic search, using sw model

(without alignments)
4152.798 Million cell updates/sec

Sequence: 1 atgaagtattcttgattc.....atgaagttgatgaagattaa 375

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

```

1:  gb.ba:*
2:  gb_hng:*
3:  gb_in:*
4:  gb_on:*
5:  gb_ov:*
6:  gb_pat:*
7:  gb_ph:*
8:  gb_pl:*
9:  gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_y1:*
15: em.ba:*
16: em.fun:*
17: em_hum:*
18: em_in:*
19: em_mn:*
20: em_om:*
21: em_ov:*
22: em_or:*
23: em_pat:*
24: em_pl:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_v1:*
30: em_hg_hum:*
31: em_hg_inv:*
32: em_hg_other:*
33: em_hg_mus:*
34: em_hg_pln:*
35: em_hg_pod:*
36: em_hg_mam:*
37: em_hg_wrt:*
38: em_sy:*
39: em_hngo_hum:*
40: em_hngo_mus:*
41: em_hngo_other:*

```

Pred. No. is the number of results predicted by chance to have a

Result No.	Score	Query Match	Length	DB	ID	Description
1	375	100.0	375	6	AX036541	AX036541 Sequence
2	48.6	13.0	189967	2	AL833802	AL833802 Mus muscu
3	48.6	13.0	206914	10	AL807815	AL807815 Mouse DNA
4	46.4	12.4	40090	3	AC006611	AC006611 Caenorhab
5	46.4	12.4	299670	2	AC006845	AC006845 Caenorhab
6	45.8	12.2	4458	2	AC115614	AC115614 Dictyoste
7	45.8	12.2	49999	6	AX015902	AX015902 Sequence
8	45.8	12.2	49999	6	AX015908	AX015908 Sequence
9	45.8	12.2	119191	9	AP004782	AP004782 Homo sapi
10	45.8	12.2	228283	10	AF312994	AF312994 Mus muscu
11	45.2	12.1	20796	3	AF216973	AF216973 Drosophil
12	45.2	12.1	87767	2	AC014497	AC014497 Drosophil
13	45.2	12.1	132000	2	AC116976	AC116976 Dictyoste
14	45.2	12.1	167062	2	AC007624	AC007624 Drosophil
15	45.2	12.1	181771	3	AC008340	AC008340 Drosophil
16	44.8	11.9	26671	3	CER11A5	283122 Caenorhabdi
17	44.4	11.8	2320	5	BC002875	BC002875 Homo sapi
18	44.4	11.8	3230	9	BC010457	BC010457 Homo sapi
19	44.4	11.8	86945	2	AC002490	AC002490 Homo sapi
20	44.4	11.8	153477	2	AC006278	AC006278 Plasmodiu
21	44.4	11.8	154470	2	AC027820	AC027820 Homo sapi
22	44.4	11.8	195770	2	AC029113	AC029113 Homo sapi
23	44.4	11.8	247175	2	AC129593	AC129593 Mus muscu
24	43.6	11.6	110000	2	PFMAL13P2-1	Continuation (2 of
25	43.6	11.6	152409	2	PFMAL13P2-1	Continuation (2 of
26	43.6	11.6	183767	2	AL807765	AL807765 Mus muscu
27	43.6	11.6	234295	2	AL714007	AL714007 Mus muscu
28	43.4	11.6	98734	2	PFMAL1P2	AL031745 Plasmodiu
29	43.4	11.6	143716	2	AL627248	AL627248 Dando rer
30	43.4	11.6	168083	2	AC092376	AC092376 Homo sapi
31	43.2	11.5	977	11	CNS061DX	AL400117 T7 end of
32	43.2	11.5	139665	2	AP000865	AP000865 Homo sapi
33	43	11.5	110000	2	PFMALAP1-1	Continuation (2 of
34	42.8	11.5	162261	2	AC119001	AC119001 Rattus no
35	42.8	11.4	142439	9	AC008749	AC008749 Homo sapi
36	42.6	11.4	59762	8	AB023032	AB023032 Arabidops
37	42.6	11.4	173016	2	AC019123	AC019123 Homo sapi
38	42.6	11.4	262428	2	AC096326	AC096326 Rattus no
39	42.4	11.3	2783	2	AF093702	AF093702 Plasmodiu
40	42.4	11.3	49144	2	AC131380	AC131380 Strongylo
41	42.4	11.3	115758	9	AC104634	AC104634 Homo sapi
42	42.4	11.3	198516	2	AC102659	AC102659 Mus muscu
43	42.4	11.3	202521	2	AL773546	AL773546 Mus muscu
44	42.4	11.3	234112	3	PFMALAP2	AL035475 Plasmodiu
45	42.2	11.3	588	8	PAB296740	AJ296740 Picea abi

ALIGNMENTS

RESULT 1
AX036541
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

Accession	Sequence	Length	Source	Accession	Sequence	Length	Source
AX036541	375 bp	DNA	linear	PAT 16-NOV-2000			
Sequence 1	from Patent	W00056885.					
AX036541							
AX036541.1	GI:11226151						

medicinal leech, hirudo medicinalis

SAWF

SAME

REFERENCE	AUTHORS	TITLE
1 (pages 1 to 375)	Hemberger, J., Scheuble, B., Strittmatter, W., Hofmann, U., Fotev, Z. and Gnessow, D.	Protein for blocking platelet adhesion

*adp. date before 1.0.1
Pub data after 1.0.1*

JOURNAL

Patent: WO 0056885-A 1 28-SEP-2000;
HEMBERGER JUERGEN (DE) ; MERCK PATENT GMBH (DE) ; SCHEUBLE BERNHARD
(DE) ; SPIITMATTER WOLFGANG (DE) ; HOTMANN UWE (DE) ; FOTEV ZISI
(DE) ; GIBSSOW DETLEF (DE)
Location/Qualifiers

FEATURES

Source

1..375

CDS

/organism="Hirudo medicinalis"
/db_xref="taxon:6421"
<64..375
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC16481.1"
/db_xref="GI:11226152"
/translation="EEREDCWFYANKRYNDPKSPKSSDLDECKKTEKTEYCTIV
FEDTVNKECYNNVDEGHELDQEFVYDENTENTLTDCBKGDAAGAGDESDVEDE
D"

BASE COUNT 123 a 62 c 85 g 105 t
ORIGIN

Query Match

100.0%; Score 375; DB 6; Length 375;
Best Local Similarity 100.0%; Pred. No. 6, 6e-86;

Matches 375; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAAGTATTTCTGATTTCTTCCCTTGGCTCGCAAGCTTGATCTCAACTACTTCT 60
DB 1 ATGAAGTATTTCTGATTTCTTCCCTTGGCTCGCAAGCTTGATCTCAACTACTTCT 60
QY 61 TCAGAACAGGTGAAGATTTGAGACGCTTTAGCCGACGAAATATACAGACTTCGAT 120
DB 61 TCAGAACAGGTGAAGATTTGAGACGCTTTAGCCGACGAAATATATACAGACTTCGAT 120
QY 121 AATCTTTTAAAGATCTCTGATCTTGAAGATGCAAAATATGTTTCAAGACGGAG 180
DB 121 AATCTTTTAAAGATCTCTGATCTTGAAGATGCAAAATATGTTTCAAGACGGAG 180
QY 181 TACGCTACATCGTTTGAAGACAGCGGTAAACAGGATGTTACTACAAATGCTTGAT 240
DB 181 TACGCTACATCGTTTGAAGACAGCGGTAAACAGGATGTTACTACAAATGCTTGAT 240
QY 241 GGTGAAGATTGACACCAAGAAATTTGTGTGACGAGAAACTCAAGGAAATTTTGG 300
DB 241 GGTGAAGATTGACACCAAGAAATTTGTGTGACGAGAAACTCAAGGAAATTTTGG 300
QY 301 ACAGACTGCGAGGTTAAAGATGCAAGTATGCGCAGAGTACAGTACAGTCAGATGAA 360
DB 301 ACAGACTGCGAGGTTAAAGATGCAAGTATGCGCAGAGTACAGTACAGTCAGATGAA 360
QY 361 GTTGATGAAGATTAA 375
DB 361 GTTGATGAAGATTAA 375

RESULT 2

AL833802

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AL833802 189967 bp DNA linear HTG 17-JUL-2002
Mus musculus chromosome 4 clone RP23-17A4, *** SEQUENCING IN
PROGRESS *** 13 unordered pieces.

AL833802.3 GI:21911787

HTG: HTGS_PHASE1.

Mus musculus

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

Plumb, B.

Submitted (16-JUL-2002) Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk

On Jul 19, 2002 this sequence version replaced gi:21738804.

----- Genome Center

FEATURES

Source

Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: BM17A4
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 186623 bases at least Q40
Consensus quality: 187480 bases at least Q30
Consensus quality: 188086 bases at least Q20
Insert size: 188767; sum-of-ctrls
Insert size: 215919; 10.9% error; agarose-fp
Quality coverage: 5.28x in Q20 bases; sum-of-ctrls Quality
coverage: 4.68x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 8454: contig of 8454 bp in length
* 8455 8554: gap of 100 bp
* 8555 19307: contig of 10753 bp in length
* 19308 19407: gap of 100 bp
* 19408 27097: contig of 7690 bp in length
* 27098 27197: gap of 100 bp
* 27198 40822: contig of 13625 bp in length
* 40823 40922: gap of 100 bp
* 40923 48761: contig of 7839 bp in length
* 48762 48861: gap of 100 bp
* 48862 94932: contig of 46071 bp in length
* 94933 95032: gap of 100 bp
* 95033 109692: contig of 14660 bp in length
* 109693 109792: gap of 100 bp
* 109793 122076: contig of 12284 bp in length
* 122077 122176: gap of 100 bp
* 122177 130769: contig of 8593 bp in length
* 130770 130869: gap of 100 bp
* 130870 146267: contig of 15398 bp in length
* 146268 146367: gap of 100 bp
* 146368 151984: contig of 5617 bp in length
* 151985 152084: gap of 100 bp
* 152085 184112: contig of 32028 bp in length
* 184113 184212: gap of 100 bp
* 184213 189967: contig of 5755 bp in length.
Location/Qualifiers
1..189967
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="4"
/clone="RP23-17A4"
/clone_1lb="RP23-17A4"
1..8454
/note="assembly-fragment:01243
fragment_chain:1
clone_end:SP6
vector_side:left"
8555..19307
/note="assembly-fragment:00590
fragment_chain:1"
19408..27097
/note="assembly-fragment:00259
fragment_chain:1"
27198..40822
/note="assembly-fragment:01468
fragment_chain:1"
40923..48761
/note="assembly-fragment:00793
fragment_chain:2"

misc_feature

misc_feature

misc_feature

misc_feature

```

misc_feature      48862..94932
                  /note="assembly-fragment:00456
                  /fragment_chain:2"
misc_feature      95033..109692
                  /note="assembly-fragment:00530
                  /fragment_chain:2"
misc_feature      109793..122076
                  /note="assembly-fragment:01290
                  /fragment_chain:2"
misc_feature      122177..130769
                  /note="assembly-fragment:01504.0"
misc_feature      130870..146267
                  /note="assembly-fragment:01504.1"
misc_feature      146368..151984
                  /note="assembly-fragment:00822
                  /fragment_chain:3"
misc_feature      152085..184112
                  /note="assembly-fragment:00663
                  /fragment_chain:3"
misc_feature      184213..189967
                  /note="assembly-fragment:00178
                  /fragment_chain:3
                  /vector_side:right"
BASE COUNT      59392 a 34302 c 33745 g 61328 t 1200 others
ORIGIN
Query Match      13.0%: Score 48.6; DB 2: Length 189967;
Best Local Similarity 48.4%: Pred. No. 0.066;
Matches 135; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

QY 97 AACAGAAATATACAGACTTCGATTAATCTTTTAAAGAGTCTTGATCTTGACGATGC 156
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 88182 AACAAAGCGCCGCAATATTGAACAGTTCTTAACAATAAAGAACTTTGGGGGAATC 88241

QY 157 AAAAAACATGTTTCAAGCGGAGTCTGCTACATCGTTTGAAGACGCGTCAACAAG 216
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 88242 ACCATTCATGACTTCAAGCTATCTACAGCAATGATGATGACAGAAAGAAAGAGAG 88301

QY 217 GAATGTTACTACATGTCGTTGATGGTGAAGAGTGTAGACCAAGAAAATTTGTTGCGAC 276
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 88302 GAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 88361

QY 277 GAAACTTCAACGAAATTTATTTGACAGACTCGGAGGTTAAAGATCAGGTATATGCGGCA 336
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 88362 GAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 88421

QY 337 GGTACAGGTGACGAGTCAAGTGAAGTGAAGATTAA 375
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 88422 GATTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 88460

RESULT 3
AL807815/c      206914 bp      DNA      linear      ROD 22-AUG-2002
LOCUS           Mouse DNA sequence from clone RP23-207H16 on chromosome 4, complete
DEFINITION
ACCESSION      AL807815
VERSION        AL807815.3 GI:22474427
KEYWORDS
SOURCE         house mouse.
ORGANISM       Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS        Leongamornlert,D.
TITLE          Submitted (22-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
JOURNAL        Cambridgehire, CB10 1SA, UK. E-mail enquiries:
               humquery@sanger.ac.uk Clone requests: clonequest@sanger.ac.uk
               On Aug 23, 2002 this sequence version replaced gi:21690116.
COMMENT
               ----- Genome Center
               Center: Wellcome Trust Sanger Institute

```

Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WormPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-207H16 is from the RPI-23 Mouse PNC Library constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACE3.6.

```

FEATURES
source
Location/Qualifiers
1..206914
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="4"
/clone="RP23-207H16"
/clone_11b="RPI-23"
BASE COUNT      64113 a 37887 c 37484 g 67430 t
ORIGIN
Query Match      13.0%: Score 48.6; DB 10: Length 206914;
Best Local Similarity 48.4%: Pred. No. 0.067;
Matches 135; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

QY 97 AACAGAAATATACAGACTTCGATTAATCTTTTAAAGAGTCTTGATCTTGACGATGC 156
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 205163 AACAAAGCGCCGCAATATTGAACAGTTCTTAACAATAAAGAACTTTGGGGGAATC 205104

QY 157 AAAAAACATGTTTCAAGCGGAGTCTGCTACATCGTTTGAAGACGCGTCAACAAG 216
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 205103 ACCATTCATGACTTCAAGCTATCTACAGCAATGATGATGACAGAAAGAAAGAGAG 205044

QY 217 GAATGTTACTACATGTCGTTGATGGTGAAGAGTGTAGACCAAGAAAATTTGTTGCGAC 276
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 205043 GAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 204984

QY 277 GAAACTTCAACGAAATTTATTTGACAGACTCGGAGGTTAAAGATCAGGTATATGCGGCA 336
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 204983 GAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 204924

QY 337 GGTACAGGTGACGAGTCAAGTGAAGTGAAGATTAA 375
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 204923 GATTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 204885

RESULT 4
AC006611      40090 bp      DNA      linear      INV 19-APR-2002
LOCUS           Caenorhabditis elegans cosmid C30F8, complete sequence.
DEFINITION
ACCESSION      AC006611
VERSION        AC006611.1 GI:4263327
KEYWORDS
SOURCE         Caenorhabditis elegans.
ORGANISM       Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
               Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE
AUTHORS        I (bases 1 to 40090)

```

200 bp overlap. Actual start of this cosmid is at base position 197 of C30F8, actual end is at 40090 of C30F8.

review of the following data: computer analysis using the program Genefinder (P. Green and L. Hillier, personal communication), the large scale EST projects of Yujii Kohara (http://www.ddbj.nig.ac.jp/cdb/elegrams/html/CE_INDEX.html) and The C elegans Oligonucleotide cloning project (<http://worfef.dcfi.harvard.edu/>), similarity to other proteins from Blastx analyses (<http://blast.wustl.edu/>), sequence conservation

using Jim Kent's MAWA alignment program (Genome Research 10:1115-1125, 2000). Individual C. elegans Genbank submissions, and personal communications with C. elegans researchers. tRNAs are predicted using the program tRNAscan-SE (Lowe, T.M. and Eddy, S.R., 1997, Nucleic Acids. Res., 25, 955-964).

gene

CDS

AUTHORS TITLE	Waterston,R. Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium Science 262 (5396), 2012-2018 (1998)
JOURNAL MEDLINE PUBMED	98069613 9831916
REFERENCE	2 (bases 1 to 40090)
AUTHORS TITLE	Latrell,P., Twyman,B. and Wilson,R. The sequence of C. elegans cosmid C30F8 Unpublished (2001)
JOURNAL REFERENCE AUTHORS TITLE	3 (bases 1 to 40090) Waterston,R.H. Direct Submission Submitted (23-FEB-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis MO 63108, USA
REFERENCE AUTHORS TITLE	4 (bases 1 to 40090) Waterston,R.H. Direct Submission Submitted (01-MAR-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis MO 63108, USA
JOURNAL	5 (bases 1 to 40090)
REFERENCE AUTHORS TITLE	Waterston,R. Direct Submission Submitted (14-AUG-2001) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA
JOURNAL	6 (bases 1 to 40090)
REFERENCE AUTHORS TITLE	Waterston,R. Direct Submission Submitted (18-OCT-2001) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA
JOURNAL	7 (bases 1 to 40090)
REFERENCE AUTHORS TITLE	Waterston,R. Direct Submission Submitted (12-FEB-2002) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA
JOURNAL	8 (bases 1 to 40090)
REFERENCE AUTHORS TITLE	Waterston,R. Direct Submission Submitted (19-APR-2002) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA
JOURNAL	Submitted by:

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality ≥ 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m3 subclone.

for a graphical representation of this cosmid sequence and its analysis see:

```
sequenceName=C3UF8;class=Sequence;
```

NEIGHBORING COSMID INFORMATION

The 5' cosmid is C46H11, 8300 bp overlap; the 3' cosmid is Y110A7A.

AUTHORS

Gloekner, G., Eichinger, L., Szafranski, K., Pachbat, J., Dear, P., Lehmann, R., Baumgart, C., Parra, G., April, J., F., Guigo, R., Kumpf, K., Tunngal, B., Cox, E., Quatl, M.A., Platzer, M., Rosenthal, A. and Noegel, A.A.

TITLE

Sequence and Analysis of Chromosome 2 of Dictyostelium

JOURNAL

The Dictyostelium Genome Sequencing Consortium

REMARK

2 (bases 1 to 4458)

AUTHORS

Baumgart, C.

TITLE

Submitted (21-MAR-2002) Genome Analysis, Institute of Molecular Biotechnology, Beutenbergstr. 11, Jena 07745, Germany

JOURNAL

is available from IMB Jena, Department of Genome Analysis (http://genome.imb-jena.de/dictyostelium/) and the University Cologne, Institute for Biochemistry I (http://www.uni-koeln.de/dictyostelium/project.shtml)

FUNDING

Agency : Deutsche Forschungsgemeinschaft (DFG).

NOTE

* This sequence will be replaced

FEATURES

* by the finished sequence as soon as it is available and * the accession number will be preserved.

SOURCE

Location/Qualifiers

ORGANISM

1. .4458

CDS

/organism="Dictyostelium discoideum"
/db_xref="taxon:44689"
/chromosome="2"
/map="5179259-5183715"
complement(544..4259)
/note="ORF_ID:dd_00748"
/pseudo
/codon_start=1

BASE COUNT 1478 a 535 c 433 g 2012 t
ORIGIN

Query Match

Best Local Similarity 12.2%; Score 45.8; DB 2; Length 4458;
Matches 89; Conservativity 55.3%; Pred. No. 0.29; Mismatches 72; Indels 0; Gaps 0;

AUTHORS

1. .4458

TITLE

Sequence and Analysis of Chromosome 2 of Dictyostelium

JOURNAL

The Dictyostelium Genome Sequencing Consortium

REMARK

2 (bases 1 to 4458)

AUTHORS

Baumgart, C.

TITLE

Submitted (21-MAR-2002) Genome Analysis, Institute of Molecular Biotechnology, Beutenbergstr. 11, Jena 07745, Germany

JOURNAL

is available from IMB Jena, Department of Genome Analysis (http://genome.imb-jena.de/dictyostelium/) and the University Cologne, Institute for Biochemistry I (http://www.uni-koeln.de/dictyostelium/project.shtml)

FUNDING

Agency : Deutsche Forschungsgemeinschaft (DFG).

NOTE

* This sequence will be replaced

FEATURES

* by the finished sequence as soon as it is available and * the accession number will be preserved.

SOURCE

Location/Qualifiers

ORGANISM

1. .4458

CDS

/organism="Dictyostelium discoideum"
/db_xref="taxon:44689"
/chromosome="2"
/map="5179259-5183715"
complement(544..4259)
/note="ORF_ID:dd_00748"
/pseudo
/codon_start=1

BASE COUNT 1478 a 535 c 433 g 2012 t
ORIGIN

source 1. .49999
/organism="Mus musculus"
/db_xref="taxon:10090"
BASE COUNT 13210 a 11814 c 10825 g 14150 t
ORIGIN

Query Match 12.2%; Score 45.8; DB 6; Length 49999;
Best Local Similarity 49.8%; Pred. No. 0.33;
Matches 116; Conservativity 0; Mismatches 117; Indels 0; Gaps 0;

AUTHORS

1. .49999

TITLE

Sequence and Analysis of Chromosome 2 of Dictyostelium

JOURNAL

The Dictyostelium Genome Sequencing Consortium

REMARK

2 (bases 1 to 4458)

AUTHORS

Baumgart, C.

TITLE

Submitted (21-MAR-2002) Genome Analysis, Institute of Molecular Biotechnology, Beutenbergstr. 11, Jena 07745, Germany

JOURNAL

is available from IMB Jena, Department of Genome Analysis (http://genome.imb-jena.de/dictyostelium/) and the University Cologne, Institute for Biochemistry I (http://www.uni-koeln.de/dictyostelium/project.shtml)

FUNDING

Agency : Deutsche Forschungsgemeinschaft (DFG).

NOTE

* This sequence will be replaced

FEATURES

* by the finished sequence as soon as it is available and * the accession number will be preserved.

SOURCE

Location/Qualifiers

ORGANISM

1. .49999

CDS

/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="2"
/map="5179259-5183715"
complement(544..4259)
/note="ORF_ID:dd_00748"
/pseudo
/codon_start=1

BASE COUNT

13135 a 11787 c 10868 g 14209 t

ORIGIN

Query Match

Best Local Similarity 12.2%; Score 45.8; DB 6; Length 49999;
Matches 116; Conservativity 0; Mismatches 117; Indels 0; Gaps 0;

AUTHORS

1. .49999

TITLE

Sequence and Analysis of Chromosome 2 of Dictyostelium

JOURNAL

The Dictyostelium Genome Sequencing Consortium

REMARK

2 (bases 1 to 4458)

AUTHORS

Baumgart, C.

TITLE

Submitted (21-MAR-2002) Genome Analysis, Institute of Molecular Biotechnology, Beutenbergstr. 11, Jena 07745, Germany

JOURNAL

is available from IMB Jena, Department of Genome Analysis (http://genome.imb-jena.de/dictyostelium/) and the University Cologne, Institute for Biochemistry I (http://www.uni-koeln.de/dictyostelium/project.shtml)

FUNDING

Agency : Deutsche Forschungsgemeinschaft (DFG).

NOTE

* This sequence will be replaced

FEATURES

* by the finished sequence as soon as it is available and * the accession number will be preserved.

SOURCE

Location/Qualifiers

ORGANISM

1. .49999

CDS

/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="2"
/map="5179259-5183715"
complement(544..4259)
/note="ORF_ID:dd_00748"
/pseudo
/codon_start=1

BASE COUNT 13135 a 11787 c 10868 g 14209 t
ORIGIN

Query Match 12.2%; Score 45.8; DB 6; Length 49999;
Best Local Similarity 49.8%; Pred. No. 0.33;
Matches 116; Conservativity 0; Mismatches 117; Indels 0; Gaps 0;

1. .49999

Sequence and Analysis of Chromosome 2 of Dictyostelium

The Dictyostelium Genome Sequencing Consortium

2 (bases 1 to 4458)

Baumgart, C.

Submitted (21-MAR-2002) Genome Analysis, Institute of Molecular Biotechnology, Beutenbergstr. 11, Jena 07745, Germany

is available from IMB Jena, Department of Genome Analysis (http://genome.imb-jena.de/dictyostelium/) and the University Cologne, Institute for Biochemistry I (http://www.uni-koeln.de/dictyostelium/project.shtml)

Agency : Deutsche Forschungsgemeinschaft (DFG).

* This sequence will be replaced

* by the finished sequence as soon as it is available and * the accession number will be preserved.

Location/Qualifiers

1. .49999

/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="2"
/map="5179259-5183715"
complement(544..4259)
/note="ORF_ID:dd_00748"
/pseudo
/codon_start=1

BASE COUNT 13135 a 11787 c 10868 g 14209 t

ORIGIN

[illegible]

FEATURES		Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany	
source		Location/Qualifiers	
		1..228283	
		/organism="Mus musculus"	
		/db_xref="taxon:10090"	
		/chromosome="1"	
		/clone="MM1"	
BASE COUNT	63688 a 49622 c 49154 g 65819 t		
ORIGIN			
Query Match	12.2%;	Score 45.6;	DB 10; Length 228283;
Best Local Similarity	49.8%;	Pred. No. 0.33;	
Matches 116; Conservative	0;	Mismatches 117;	Indels 0; Gaps 0;
QY	143 ATCTTGACGATCCAAAAACATGTTTCAGACGAGTACTGCTACATGCTTTTGAG	202	
Db 148716	ATTTTGAAGAAAAAAATCTATATTTACAGATTCACACTGTTATATATGAGGAAG	148775	
QY	203 ACACGGTCACCAAGGATGTTACTACATATGCTTGATGTTGAAGACTAGACCAAGAA	262	
Db 148776	AAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	148935	
QY	263 AATTTGTTTCGACGAAACTTACGCGAAATTTATTGACAGACTCGAGGGTTAAAGATG	322	
Db 148936	AAGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	148955	
QY	323 CAGGTAAATGGCGGAGTACAGCTGACGAGCTCAGATGAGATTGATGAAGATTAA	375	
Db 148996	AAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	148948	

[illegible]

[illegible]

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 25, 2003, 01:11:05 ; Search time 38 Seconds

(without alignments)
2629.946 Million cell updates/sec

Title: US-09-936-737A-1

Perfect score: 674
Sequence: 1 atgaagtattcttgcatttc.....atgaagtgtgaagaattaa 375

Scoring table:

BIOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL-frame+ n2p.model -DEV-xlh
-O/cgn2.1/USPTO_spool/US0936737/runat_14022003_100557_1379/app_query.fasta_1.519
-DB-A-Geneseq_101002 -OPMT-fastan -SUFFIX-frag -MIMATCH=0.1 -LOOPCL=0
-LOOEXT=0 -UNITS-bits -STRAT=1 -END=-1 -MATRIX-biosum62 -TRANS-human40.cdi
-LIST=45 -DOCATIGN=200 -THR SCORE-pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE-LOCAL -OUTFMT-plt -NORM-ext -HAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USR=US0936737.qcgn2.1_1.22.qrunat_14022003_100557_1379 -NCP=6 -ICPU=3
-NO_XLPHY -NO_MAMP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV-TIMEOUT=120
-WARN_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_101002.*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	573	85.0	103	AA94746	Platelet binding 1
2	88.5	13.1	147	AA22950	Leech antiplatelet
3	82.5	12.2	368	AB86867	Drosophila melanog
4	82.5	12.2	412	AA668010	S. frugiperda immu
5	79.5	11.8	150	AA95228	Mouse soluble BMP
6	79.5	11.8	150	AA96200	Muscle bone morph
7	79.5	11.8	200	AA95234	Mouse incomplete B
8	79.5	11.8	1038	AA95224	Mouse BMP receptor
9	79.5	11.8	1038	AA96198	Bone morphogenetic
10	79.5	11.8	1038	AA96247	Bone BMP receptor
11	79.5	11.2	323	AA958349	Arabidopsis thalia
12	75.5	11.2	331	AA958348	Arabidopsis thalia
13	75.5	11.2	336	AA958347	Arabidopsis thalia
14	75.5	11.2	418	AA94260	Mouse ice-4 protel
15	75.5	11.2	419	AA950109	Murine caspase-12
16	75.5	11.2	419	AA94259	Mouse ice-4 protel
17	75.5	11.2	419	AA929108	Murine caspase-12
18	75.5	11.2	432	AA96769	Human interleukin-
19	75.5	11.2	432	AA96464	Murine Ich-3. Mus
20	75.5	11.2	719	AB927652	Streptococcus poly
21	74.5	11.1	482	AB950229	Human transcriptio
22	74.5	11.4	1547	AA921168	Human TRIC-12 pro
23	74	11.0	248	AB966151	Drosophila melanog
24	74	11.0	577	AA930586	Arabidopsis thalia
25	74	11.0	578	AA930585	Arabidopsis thalia
26	74	11.0	612	AA930584	Arabidopsis thalia
27	74	11.0	1481	AA903740	rchd528 gene produ
28	74	11.4	1481	AA95160	Amino acid sequenc
29	74	11.4	1481	AA98299	Human rchd528 gene
30	74	11.4	1481	AA91926	Amino acid sequenc
31	74	11.4	1481	AA978508	Richd528 amino acid
32	74	11.4	1481	AA968447	Amino acid sequenc
33	73.5	10.9	228	AA935750	Endochitinase Ctr52
34	73.5	10.9	562	AA944001	Drosophila melanog
35	73.5	10.9	1154	AB966151	Drosophila melanog
36	72.5	10.8	660	AB966151	Drosophila melanog
37	72.5	10.8	976	AA976012	Human poly(ADP-rib
38	72.5	10.8	976	AA976012	Human poly(ADP-rib
39	72	11.1	114	AA934450	Zea mays protein f
40	72	10.7	143	AA920404	Arabidopsis thalia
41	72	10.7	145	AA920403	Arabidopsis thalia
42	72	10.7	587	AA910554	S. cerevisiae aspa
43	72	10.7	1092	AA979135	Human protein Seq
44	72	10.7	1092	AA940100	Human polypeptide
45	72	10.7	1094	AA980119	Human protein Seq

ALIGNMENTS

RESULT 1
AA94746
ID AA94746 standard: Protein; 103 AA.

AC AA94746;

DT 29-JAN-2001 (first entry)

DE Platelet binding inhibitor protein Saratin amino acid sequence.

KW Saratin: medicinal leech; thromboembolic disease; intraocular lens;

KW collagen-dependent platelet adhesion inhibition;

KW posterior capsule opacification.

OS Hirudo medicinalis.

XX
PN WO200056885-A1.
XX
PD 28-SEP-2000.

PN	US5861498-A.	
XX		
PD	19-JAN-1999.	
XX		
PF	31-OCT-1996;	96US-0741134.
XX		
PR	01-NOV-1995;	95US-0007163.
PR	31-OCT-1996;	96US-0741134.
XX		
PA	(UYJE-) UNIV JEFFERSON THOMAS.	
XX		

Alnemri ES, Fernandes-Alnemri T, Litwack G,
PI
XXX
WPI; 1999-130433/11.
N-PSDB; AAx04441.
xx

Isolated immunophilin FKBP46 nucleic acids - useful for developing products for the study and identification of immunosuppressive agents for treating e.g. transplant and tissue graft patients

This sequence represents the immunophilin FKBP46 from the moth *Spodoptera frugiperda*. FKBP46 is a novel insect cell nuclear immunophilin, which can be used to study and identify additional immunosuppressive drugs that bind to it. Such immunosuppressive agents can be used in treating transplant and tissue graft patients.

Q Sequence 412 AA;

Alignment Scores:	
Seed. No.:	
Score:	1.11
Percent Similarity:	82.50
Percent Local Similarity:	41.74%
Query Match:	27.63%
Gaps:	12.24%
	20
Length:	412
Matches:	32
Conservative:	16
Mismatches:	40
Indels:	27
Gaps:	5

S-09-936-737A-1 (1-375) X AAW68010 (1-412)

43 CTGATTCAACTACTTCTTGAGAGAAGCGTAGACATTGTGGACGTTTACCAGACA 102
|||::: :: :: :: |||:::
116 LeuValProAlaLysAsnLysArgLysLeu(LunSm)Lan(SnpR)AthrLalaSnLys 135

103 AATATACAGACTTCGATAAATCTTTTAGAGAGTCC-----TGTGATCTT 147

||| |||
||||:
|||:

136 lysalalysprosplyslysalacglylsasnserralaprolaalagluseraspser 155
148 GACGAATGCAAAAAACATCTGTTCATCAGCGAATGGTCTTTCTG

```

|||||
156 AspAspAsp-----GTTGGAGACACG 207

```

GluAspGln 162

208 GTCAACAAGGAATGGTAACTACAAATCCTCCATCAATCATTAATTTAATTGTTT

[illegible]

-----PhelouspcglwaspIleaspthraspcglu--- 176

QY	268	GTGTGCGAAACCTTACCGAAATATTATTGACACATCGCAGGTAAAGATGACAGT	327
Db	177	--AsnAspIuSerPheIuMetAsn-----ThSerAlaGluIuAspAspSerAsp	193
QY	328	AATGCGCAGGATACAGGAGCAGTACAGTACAGTACAGTATGTAAGAT	372
Db	194	GIuGIuAspAspAspGluAspGluGIuAspGluIuAspGluIuAspAsp	208

ID AAR05228
 RESULT 5
 AAR95228

12	AA95228	Standard; 150 AA
XX		
AC	AA95228;	

XX
DT 31-DEC-1996 (first entry)
XX

Mouse soluble BMP receptor kinase protein-3 (BRK-3).

KW BMP type II receptor kinase-3; BRK-3; bone morphogenetic protein,
KW BMP type I receptor kinase; BMP receptor.
XX

mus sp.
XX

PN W09614579-A1.
XX
PD 17-MAY-1996

XX 1 / MAY-1990.
PF 30-OCT-1995; 95WO-US14027

05-JUN-1995; 95US-0462467.
04-NOV-1984; SATC 000447.

NOV 1954; 9405-0334178.
XX
PA
(PROC) PROCTER & GAMBLE CO.

Rosenbaum JS;

WPI; 1996-251887/25.
N-PSDB; AAT28024.

Assays for bone morphogenetic protein activities - using complex of BMP type I receptor and

BRK-3
X
receptor kinase protein and BMP receptor kinase protein

Disclosure; Page 63; 101pp; English.

Mouse soluble bone morphogenetic protein (BMP) type II receptor kinase protein-3 (BRK-3) (AAR952228) lacks the regions of the

length receptor (AAR955224) not required for BMP binding, and is the product of a truncated CDNA sequence (AAT28023). A BMP receptor kinase protein complex formed

receptor kinase proteins (see also AAR5222-27 and AAR5230-32).

useful for screening cpds. for BMP receptor affinity or for determining the concentration of a BMP receptor ligand in a clinical sample. The complex can be used as a

with vectors carrying the appropriate DNA sequences (see also AAT28018-30).

Sequence 150 AA;

Assignment Scores:
Assigned. No.: 1 84

Score:	1.04	Length:	150
Percent Similarity:	79.50	Matches:	27
	40.46%	Conservative:	26

Local Similarity:	20.61%	Mismatches:	35
every Match:	11.80%	Indels:	43
	17		2

Gaps

03-930-13/A-1 (1-375) X AAR95228 (1-150)

10 TTCTGATTTCTCTCTTGCCTCGACGTTGCTGATCTACACTACTCTTCTCAGAAGA 69
|||

```

Db      9 PheargValProtrpleuLeutrPalValleuValSerThrThraAlaSerGln 28
QY      70 CGTGAAGATGTTGGACGTTTACCGAAGAAATATACAGACTTCGATAAATCTTT 129
      29 AsngInGln-----ArgLeuCysAlaPheLysAspProtyr 40
      130 AACAAGTCTCTGACTTCACGATGCAAAAAACA-----TGT 168
Db      41 GlnGlnAspLeuGlyLeuGlyLeuSerArgIleSerHisGluAsnGlyThrIleLeuGly 60
QY      169 TTCAGACGAGACTGCTACATCGTTTGA-----GACACGGTC 210
      61 SerLysGlySerThrCysTyrGlyLeuTrpGluLysSerLysGlyAspIleAsnLeuVal 80
QY      211 AACAAGAAATGTTACTACATGCGTTCGTAAGAGTTAGACCAAGAAATTTGTT 270
      81 LysGlnGlyCysTrpSerHisIleGlyAspProGlnGluCysHisTyrGlnGluCysVal 100
QY      271 GTC----- 273
Db      101 ValThrThrThrProProSerIleGlnAsnGlyThrTyrArgPheCysCysSerThr 120
QY      274 -----GACGAAAACTTCACGCAAAATAT 297
      121 AspLeuCysAsnValAsnPheThrGluAsnPhe 131
RESULT 6
AAR96200
ID      AAR96200 standard; Protein; 150 AA.
XX
AC      AAR96200;
XX
DE      22-AUG-1996 (first entry)
XX
DE      Soluble bone morphogenetic protein receptor kinase-3 fragment.
XX
KW      Mouse; bone morphogenetic protein receptor kinase-3; soluble;
KW      fibroblast; cysteine box; COS; CHO; cell culture; antibody;
KW      diagnostic; bone disorder; osteogenic; drug screening;
KW      bone morphogenetic protein-agonist; reporter gene;
KW      bone morphogenetic protein-antagonist; hormone-responsive element.
XX
OS      Mus musculus.
XX
FH      Key Location/Qualifiers
FT      Peptide 116..123
      /note="Cysteine box"
XX
XX      W09614412-A2.
XX      17-MAY-1996.
XX      30-OCT-1995; 95WO-US14085.
XX      04-NOV-1994; 94US-0334179.
XX      (PROC ) PROCTER & GAMBLE CO.
XX      PA
XX      PI Nohno T, Rosenbaum JS;
XX      DR WPI. 1996-251762/25.
XX      N-PSDB; AAT27227.
XX
PT      Isolated bone morphogenic protein receptor kinase protein - used to
PT      determine if a test cpd. is capable of binding to, or is
PT      (ant)agonist of BMP receptor kinase protein transcription
XX
XX      Claim 4; Page 60; 87pp; English.
XX
XX      The sequence represents a soluble fragment (extracellular domain)
XX      of mouse bone morphogenetic protein (BMP) receptor type-II kinase-3
XX      (BRK-3), which induces cellular differentiation in response to BMP.
XX      A gene encoding full-length BRK-3 (AAR96198) may be isolated from

```

```

CC      NIH3T3 mouse embryo fibroblast cDNA using PCR. The soluble fragment
CC      binds BMP, and has all regions not required for BMP binding
CC      deleted. Like other transforming growth factor-beta superfamily
CC      receptors, the extracellular domain of BRK-3 contains a single
CC      cysteine box. The BRK-3 receptor gene may be inserted in a vector
CC      and expressed in a CHO or COS cell culture. The receptor and
CC      antibodies against it may be used in diagnostic assays for BMP
CC      disorders, or in therapy to bind or scavenge BMPs. In addition,
CC      expression of the BRK-3 gene along with a reporter gene under the
CC      control of a hormone-responsive element in a cell culture may be
CC      used to screen compounds for BRK-agonist or -antagonist activity,
CC      by monitoring reporter gene expression.
XX
SQ      Sequence 150 AA.
XX
Alignment Scores:
Pred. No.: 1.84 Length: 150
Score: 79.50 Matches: 27
Percent Similarity: 40.46% Conservative: 26
Best Local Similarity: 20.61% Mismatches: 35
Query Match: 11.80% Indels: 43
DB: 17 Gaps: 4
XX
US-09-936-737A-1 (1-375) x AAR96200 (1-150)
QY      10 TTCCTGATTCCTCCCTTCCTGCGACGCTTCGATCTCACTCTTCAGAGAA 69
      9 PheargValProtrpleuLeutrPalValleuValSerThrThraAlaSerGln 28
QY      70 CGTGAAGATGTTGGACGTTTACCGAAGAAATATACAGACTTCGATAAATCTTT 129
      29 AsngInGln-----ArgLeuCysAlaPheLysAspProtyr 40
QY      130 AACAAGTCTCTGACTTCACGATGCAAAAAACA-----TGT 168
      41 GlnGlnAspLeuGlyLeuGlyLeuSerArgIleSerHisGluAsnGlyThrIleLeuGly 60
QY      169 TTCAGACGAGACTGCTACATCGTTTGA-----GACACGGTC 210
      61 SerLysGlySerThrCysTyrGlyLeuTrpGluLysSerLysGlyAspIleAsnLeuVal 80
QY      211 AACAAGAAATGTTACTACATGCGTTCGTAAGAGTTAGACCAAGAAATTTGTT 270
      81 LysGlnGlyCysTrpSerHisIleGlyAspProGlnGluCysHisTyrGlnGluCysVal 100
QY      271 GTC----- 273
Db      101 ValThrThrThrProProSerIleGlnAsnGlyThrTyrArgPheCysCysSerThr 120
QY      274 -----GACGAAAACTTCACGCAAAATAT 297
      121 AspLeuCysAsnValAsnPheThrGluAsnPhe 131
RESULT 7
AAR95234
ID      AAR95234 standard; Protein; 200 AA.
XX
AC      AAR95234;
XX
DE      31-DEC-1996 (first entry)
XX
DE      Mouse incomplete BMP receptor kinase protein-3 (BRK-3).
XX
KW      BMP type II receptor kinase-3; BRK-3; bone morphogenetic protein;
KW      BMP type I receptor kinase; BMP receptor.
XX
OS      Mus sp.
XX
FH      Key Location/Qualifiers
FT      Domain 151..172
      /label="Transmembrane_domain"
XX
XX      W09614579-A1.

```

```

XX 17-MAY-1996.
PD 30-OCT-1995; 95WO-US14027.
XX 05-JUN-1995; 95US-0462467.
XX 04-NOV-1994; 94US-0334178.
XX (PROC ) PROCTER & GAMBLE CO.
XX Rosenbaum JS;
XX WPI: 1996-251887/25.
XX N-PSDB: AAT28030.
XX Assays for bone morphogenetic protein activities - using complex of
XX BMP type I receptor kinase protein and BMP receptor kinase protein
XX BRK-3
XX Disclousure; Page 80; 101pp; English.
XX Mouse incomplete bone morphogenetic protein (BMP) type II receptor
XX transmembrane and intracellular juxtamembrane region of full-length
XX BRK-3 (AAR95224) but is incapable of signalling due to deletion of the
XX intracellular kinase domain. It is the product of a truncated cDNA
XX sequence (AAT28030). A BMP receptor kinase protein complex formed of
XX full-length, truncated, incomplete or soluble BRK-3 and full-length,
XX incomplete or soluble BMP type I receptor kinase proteins (see also
XX AAR95222-33) is useful for screening cpds. for BMP receptor affinity
XX or for determining the concentration of a BMP receptor ligand in a
XX clinical sample. The complex can be expressed by host cells
XX co-transfected with vectors carrying the appropriate DNA sequences
XX (see also AAT28018-30).
XX
XX Sequence 200 AA:
XX
Alignment Scores:
Pred. No.: 1.99 Length: 200
Score: 79.50 Matches: 27
Percent Similarity: 40.46% Conservative: 26
Best Local Similarity: 20.61% Mismatches: 35
Query Match: 11.80% Indels: 43
DB: 17 Gaps: 4
US-09-936-737a-1 (1-375) x AAR95234 (1-200)
QY 10 TTCTGATTTCCTTCCTTCCTCGCAAGCTTGATCTCACTCTTTCAGAGAA 69
DB 9 PheAlGValPProTrpLeuLeuTrpAlaValLeuLeuValSerThrAlaAlaSerGln 28
QY 70 CGTGAAGATTGTTGGACGTTTACGCCGAAAGAAATATACAGACTTCGATAATCTTTT 129
DB 29 AangInguL-----ArgLeuCySaIaPheLysAspProTyr 40
QY 130 AAGAAGTCCTCGATCTTGACGAATGCAAAAAAACA-----TCG 168
DB 41 GInGInAspLeuGLYllleGlyLysSerArgLleSerHISgluAsnGlyThrIleLeuCyS 60
QY 169 TTCAGACGGAGGACTGCTACATCGTTTTCGA-----GACACGGTC 210
DB 61 SerLysGlySerThrCysTyrGlyLeuTrpLysSerLysGlyAspIleAsnLeuVal 80
QY 211 AACAGGAATGTACTACATGCTGTTAGGTGTAAGAGTTAGACCAAGAAATTTGTT 270
DB 81 LysGInGlyCysTrpSerHisIleGlyAspProGlnInGlyShisTyrGlnGlyVal 100
QY 271 GTC----- 273
DB 101 ValThrThrThrProProSerIleGlnAsnGlyThrTyrArgPheCysCysSerThr 120
QY 274 -----GACGAAACCTTCACGGAATAATTAAT 297
::: |||||

```

```

DB 121 AspleuCySaAsnValAsnPheThrGluAsnPhe 131
RESULT 8
AAR95224
ID AAR95224 standard; Protein; 1038 AA.
XX AAR95224;
XX 31-DEC-1996 (first entry)
XX Mouse BMP receptor kinase protein-3 (BRK-3).
XX
XX BMP type II receptor kinase-3; BRK-3; bone morphogenetic protein;
XX BMP type I receptor kinase; BMP receptor.
XX Mus sp.
XX
XX Key Location/Qualifiers
XX Domain 151..172
XX /Label= Transmembrane_domain
XX
XX W09614579-A1.
XX 17-MAY-1996.
XX
XX 30-OCT-1995; 95WO-US14027.
XX
XX 05-JUN-1995; 95US-0462467.
XX 04-NOV-1994; 94US-0334178.
XX (PROC ) PROCTER & GAMBLE CO.
XX
XX Rosenbaum JS;
XX
XX WPI: 1996-251887/25.
XX N-PSDB: AAT28020.
XX
XX Assays for bone morphogenetic protein activities - using complex of
XX BMP type I receptor kinase protein and BMP receptor kinase protein
XX BRK-3
XX Claim 2; Page 61-62; 101pp; English.
XX
XX Full-length mouse bone morphogenetic protein (BMP) type II receptor
XX kinase protein-3 (BRK-3) (AAR95224) is a receptor capable of binding
XX BMP and transducing a signal initiated by the binding. Its amino
XX acid sequence was deduced from a cDNA clone (AAT28020) obtd. from
XX NIH3T3 fibroblasts. A BMP receptor kinase protein complex formed
XX of full-length, truncated, incomplete or soluble BRK-3 and full-
XX length, incomplete or soluble BMP type I receptor kinase proteins
XX (see also AAR95222-23 and AAR95225-34) is useful for screening cpds.
XX for BMP receptor affinity or for determining the concentration of a
XX BMP receptor ligand in a clinical sample. The complex can be
XX expressed by host cells co-transfected with vectors carrying the
XX appropriate DNA sequences (see also AAT28018-30).
XX
XX Sequence 1038 AA:
XX
Alignment Scores:
Pred. No.: 3.19 Length: 1038
Score: 79.50 Matches: 27
Percent Similarity: 40.46% Conservative: 26
Best Local Similarity: 20.61% Mismatches: 35
Query Match: 11.80% Indels: 43
DB: 17 Gaps: 4
US-09-936-737a-1 (1-375) x AAR95224 (1-1038)
QY 10 TTCTGATTTCCTTCCTTCCTCGCAAGCTTGATCTCACTCTTTCAGAGAA 69
DB 9 PheAlGValPProTrpLeuLeuTrpAlaValLeuLeuValSerThrAlaAlaSerGln 28
QY 70 CGTGAAGATTGTTGGACGTTTACGCCGAAAGAAATATATACAGACTTCGATAATCTTTT 129

```

```

Db      29 AsnGlnGlu-----ArgLeuCysAlaPheLysAspProTyr 40
QY      130 AAGAAGTCCTGTGATCTTGACGAATGCAAAAAACA-----TGT 168
Db      41 GlnGlnAspLeuGlyIleGlyIleSerArgIleSerHisGlnAsnGlyThrIleLeuCys 60
QY      169 TTCAGACGAGAGTACTGCTACATCGTTTGA-----GACACGGTC 210
Db      61 SerLysGlySerThrCysTyrGlyLeuTrpGlnLysSerLysGlyAspIleAsnLeuVal 80
QY      211 AACAGAGATGTTACTACATGCTGATGAGGAGGAGGTTACACCAAGAAAAATTGTT 270
Db      81 LysGlnIlyCysTrpSerHisIleGlyAspProGlnIlyCysHisTyrGlnIlyCysVal 100
QY      271 GTC----- 273
Db      101 ValThrThrThrProProSerIleGlnAsnGlyThrTyrArpPheCysCysSerThr 120
QY      274 -----GACGAAAACTTCACGGAATAATTAT 297
Db      121 AspleuCysAsnValAsnPheThrGluAsnPhe 131

RESULT 9
ID      AAR96198 standard; Protein; 1038 AA.
XX
XX      AAR96198;
AC
DT      22-AUG-1996 (first entry)
XX
DE      Bone morphogenetic protein type-II receptor kinase-3.
XX
KW      Mouse; bone morphogenetic protein receptor kinase-3; fibroblast;
KW      cysteine box; protein-kinase domain; COS; CHO; antibody;
KW      cell culture; diagnostic; bone disorder; osteogenic;
KW      bone morphogenetic protein-agonist; drug screening; reporter gene;
KW      bone morphogenetic protein-antagonist; hormone-responsive element.
XX
OS      Mus musculus.
XX
FH      Key
FH      Region
FT      Location/Qualifiers
FT      1..150
FT      /note= "Soluble BRK-3 fragment (AAR96199, claim 4)"
FT      Modified-site
FT      53..57
FT      /note= "N-glycosylation site"
FT      Modified-site
FT      110..112
FT      /note= "N-glycosylation site"
FT      Peptide
FT      116..123
FT      /note= "Cysteine box"
FT      Domain
FT      151..172
FT      /note= "Transmembrane region"
FT      Domain
FT      173..1038
FT      /note= "Intracellular domain"
FT      Domain
FT      200..504
FT      /note= "Protein-kinase domain"
FT      Region
FT      505..1038
FT      /note= "Long C-terminal region"
XX
XX      WO9614412-A2.
XX
XX      17-MAY-1996.
XX
XX      30-OCT-1995; 95WO-US14085.
XX
XX      04-NOV-1994; 94US-0334179.
XX
XX      (PROC ) PROCTER & GAMBLE CO.
XX
XX      Nohno T, Rosenbaum JS;
XX
XX      WPI; 1996-251762/25.
XX
XX      N-PSDB; AAT27225.

```

```

XX      Isolated bone morphogenic protein receptor kinase protein - used to
XX      determine if a test cpd. is capable of binding to, or is
XX      (ant)agonist of BMP receptor kinase protein transcription
XX      Claim 1; Page 56-59; 87pp; English.
XX
XX      The sequence represents mouse bone morphogenetic protein (BMP)
XX      receptor type-II kinase-3 (BRK-3), which induces cellular
XX      differentiation in response to BMP. A gene encoding BRK-3 may be
XX      isolated from NIH3T3 mouse embryo fibroblast cDNA using PCR. The
XX      sequence of a soluble fragment is given in AAR96200. The
XX      extracellular domain of BRK-3 contains a single cysteine box
XX      located near the transmembrane region. The predicted cytoplasmic
XX      region contains all of the consensus sequences characterizing a
XX      protein-kinase domain with predicted specificity for serine and
XX      threonine residues. The kinase domain is followed by an extremely
XX      long C-terminus, making the intracellular domain much larger than
XX      that of any other receptor in the transforming growth factor-beta
XX      receptor family. The BRK-3 receptor gene may be inserted in a
XX      vector and expressed in a CHO or COS cell culture. The receptor
XX      and antibodies against it may be used in diagnostic assays for BMP
XX      disorders, or in therapy to bind or scavenge BMPs. In addition,
XX      expression of the BRK-3 gene along with a reporter gene under the
XX      control of a hormone-responsive element in a cell culture may be
XX      used to screen compounds for BRK-agonist or -antagonist activity,
XX      by monitoring reporter gene expression.
XX
XX      SQ      Sequence      1038 AA;
XX
XX      Alignment Scores:
XX      Pred. No.:      3.19      Length:      1038
XX      Score:      79.50      Matches:      27
XX      Percent Similarity:      40.46%      Conservative:      26
XX      Best Local Similarity:      20.61%      Mismatches:      35
XX      Query Match:      11.80%      Indels:      4
XX      DB:      17      Gaps:      4
XX
XX      US-09-936-737a-1 (1-375) x AAR96198 (1-1038)
QY      10 TTCTGATTTCCTCTTCCTTCCTTCGCGACGCTGCTGATCTCAACTCTTCACAGAA 69
Db      9 PheArgValProTrpLeuLeuTrpAlaValLeuLeuValSerThrThrAlaIleSerGln 28
QY      70 CGTGAAGATTGTTGGACGTTTTCACGCAACAGAAAAATATACAGACTTCGATAATCTT 129
Db      29 AsnGlnGlu-----ArgLeuCysAlaPheLysAspProTyr 40
QY      130 AAGAAGTCCTGTGATCTTGACGAATGCAAAAAACA-----TGT 168
Db      41 GlnGlnAspLeuGlyIleGlyIleSerArgIleSerHisGlnAsnGlyThrIleLeuCys 60
QY      169 TTCAGACGAGAGTACTGCTACATCGTTTGA-----GACACGGTC 210
Db      61 SerLysGlySerThrCysTyrGlyLeuTrpGlnLysSerLysGlyAspIleAsnLeuVal 80
QY      211 AACAGAGATGTTACTACATGCTGATGAGGAGGAGGTTACACCAAGAAAAATTGTT 270
Db      81 LysGlnIlyCysTrpSerHisIleGlyAspProGlnIlyCysHisTyrGlnIlyCysVal 100
QY      271 GTC----- 273
Db      101 ValThrThrThrProProSerIleGlnAsnGlyThrTyrArpPheCysCysSerThr 120
QY      274 -----GACGAAAACTTCACGGAATAATTAT 297
Db      121 AspleuCysAsnValAsnPheThrGluAsnPhe 131

RESULT 10
ID      AAR96247
XX
XX      AAR96247 standard; Protein; 1038 AA.
AC

```

CC This represents a mouse BMP receptor kinase protein (BRK)-3. The mouse
CC BRK-1 and BRK-2 protein sequences (AA0865248 and AA0865249) can be used in
CC the method of the invention of determining whether a compound can bind
CC to a bone morphogenetic protein (BMP) receptor kinase protein complex.
CC The method comprises allowing a compound in a sample to bind to the
CC complex, where the complex is comprised of (i) a BMP; (ii) a BRK protein
CC (iii) an ACRIIRB receptor. The method can be used to determine the
CC concentration of a BMP receptor ligand in a sample by comparing the
CC binding to a standard curve prepared with known concentrations of BMP
CC ligand. The method can also be used to determine whether a test compound
CC produces a signal on binding to a BMP receptor protein complex. The
CC method is useful for determining whether a ligand, such as a known or
CC putative drug, can bind to and/or activate the receptors.
XX
XX Sequence 1038 AA;

US-09-936-737A-1 (1-375) x AAW86247 (1-1038)

QY	10	TTCTGATTCCTTCCTTGGCTGCGAAGCTGGTGATCTCAACTCTCTTCACAGAA	63
	11	:::: :::: ::: ::: :::	
Db	9	PheArgValProTyrPheLeuThrPheValIleuValSerThrThrIalaIaSerGln	28
QY	70	CGTGAAGATTGTGGACGCTTTTACGCGAACAGAAATATACAGACTTCGATTAATCTTT	122
	71	::::: :::: :::: :::: ::::	
Db	29	AsnGlnGlu-----ArgLeuGlnAlaPheLysAspProTyr	40
QY	130	AAGAACTCCTCGATTCTTGACGAATGCAAAAAA-----TGT	166
	131	::::: :::: :::: :::: :::: :::: :::: :::: ::::	
Db	41	GlnGlnAspLeuGlyIleGlyGlnSerArgIleSerHisGlnAsnGlyThrIleLeuGln	60
QY	169	TTCAAGACGCGAGTACTGCTACATCGCTTTTGGAA-----GACACGGCTC	210
	170	:::: ::::	
Db	61	SerTyrGlySerThrTyrGlyIleuThrGlnLysSerTyrGlyAspIleAsnLeuVal	80

Qy	211	AAACAGAAAGTGTACATGCATGCTGATGGTGAAGAGTGTACACCAAGAAATTTGTT	270
		::: ::: ::::: :::::	
Db	81	LyGIcInGIcYcStrIpSerHIStIEGIcYsPrOGInGIcYsHIStYrGIcInGIcYsVal	100
Qy	271	GTC-----	273
Db	101	ValtThrThrProSeRIeGInaEnGIcYrThrYrArgPheCysCysSerThr	120
Qy	274	-----GACGAAGACTTCACGGAATTAAT	297
		::: :::	
Db	121	AspLeuCysAsnValAsnPhetHrGIcUasPhe	131
RESULT 11			
ID	AA658349		
XX	AA658349 standard; Protein; 323 AA.		
XX	AA658349;		
XX	18-OCT-2000 (first entry)		
XX	Arabidopsis thaliana protein fragment SEQ ID NO: 75310.		
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 75310.		
XX	Protein identification; signal transduction pathway; metabolic pathway;		
KW	hybridisation assay; genetic mapping; gene expression control; promoter;		
KW	termination sequence.		
XX	Arabidopsis thaliana.		
OS	Arabidopsis thaliana.		
XX	EP1033405-A2.		
XX	06-SEP-2000.		
XX	25-FEB-2000; 2000EP-0301439.		
XX	25-FEB-1999; 99US-0121825.		
PR	05-MAR-1999; 99US-0123180.		
PR	09-MAR-1999; 99US-0123548.		
PR	23-MAR-1999; 99US-0125788.		
PR	25-MAR-1999; 99US-0126264.		
PR	29-MAR-1999; 99US-0126785.		
PR	01-APR-1999; 99US-0127462.		
PR	06-APR-1999; 99US-0128234.		
PR	08-APR-1999; 99US-0128714.		
PR	16-APR-1999; 99US-0129845.		
PR	19-APR-1999; 99US-0130077.		
PR	21-APR-1999; 99US-0130449.		
PR	23-APR-1999; 99US-0130510.		
PR	28-APR-1999; 99US-0130891.		
PR	30-APR-1999; 99US-0132048.		
PR	30-APR-1999; 99US-0132407.		
PR	04-MAY-1999; 99US-0132464.		
PR	05-MAY-1999; 99US-0132485.		
PR	06-MAY-1999; 99US-0132486.		
PR	07-MAY-1999; 99US-0132487.		
PR	11-MAY-1999; 99US-0132863.		
PR	11-MAY-1999; 99US-0134256.		
PR	14-MAY-1999; 99US-0134218.		
PR	14-MAY-1999; 99US-0134219.		
PR	14-MAY-1999; 99US-0134219.		
PR	14-MAY-1999; 99US-0134221.		
PR	14-MAY-1999; 99US-0134370.		
PR	18-MAY-1999; 99US-0134768.		
PR	19-MAY-1999; 99US-0134941.		
PR	20-MAY-1999; 99US-0135124.		
PR	21-MAY-1999; 99US-0135353.		
PR	24-MAY-1999; 99US-0135629.		
PR	25-MAY-1999; 99US-0136021.		
PR	27-MAY-1999; 99US-0136392.		
PR	28-MAY-1999; 99US-0136782.		
PR	01-JUN-1999; 99US-0137222.		
PR	03-JUN-1999; 99US-0137528.		
PR	04-JUN-1999; 99US-0137502.		
PR	07-JUN-1999; 99US-0137724.		

PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139753.
 PR 18-JUN-1999; 99US-0139817.
 PR 21-JUN-1999; 99US-0139859.
 PR 22-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 28-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 30-JUN-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144632.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 23-JUL-1999; 99US-0145226.
 PR 26-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147300.
 PR 06-AUG-1999; 99US-0147410.

PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149920.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150568.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151303.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 13-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0158293.
 PR 13-OCT-1999; 99US-0159294.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161922.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Alignment Scores:
 Pred. No.:
 Score:
 Percent Similarity:

6.54
 75.50
 34.758

Length: 323
 Matches: 34
 Conservative: 15

	PR	14-OCT-1999;	99US-0159638.	
	PR	18-OCT-1999;	99US-0159584.	
	PR	21-OCT-1999;	99US-0160741.	
	PR	21-OCT-1999;	99US-0160767.	
	PR	21-OCT-1999;	99US-0160768.	
	PR	21-OCT-1999;	99US-0160770.	
	PR	21-OCT-1999;	99US-0160814.	
	PR	21-OCT-1999;	99US-0160815.	
	PR	22-OCT-1999;	99US-0160980.	
	PR	22-OCT-1999;	99US-0160981.	
	PR	22-OCT-1999;	99US-0160989.	
	PR	23-OCT-1999;	99US-0161404.	
	PR	23-OCT-1999;	99US-0161405.	
	PR	25-OCT-1999;	99US-0161406.	
	PR	26-OCT-1999;	99US-0161359.	
	PR	26-OCT-1999;	99US-0161360.	
	PR	26-OCT-1999;	99US-0161361.	
	PR	26-OCT-1999;	99US-0161920.	
	PR	28-OCT-1999;	99US-0161920.	
	PR	28-OCT-1999;	99US-0161992.	
	PR	28-OCT-1999;	99US-0161993.	
	PR	29-OCT-1999;	99US-0162142.	

Alignment Scores:		Pred. No.:	6.58	Length:	331
Score:			75.50	Matches:	34
Percent Similarity:			34.75%	Conservative:	15
Best Local Similarity:			24.11%	Mismatches:	41
Query Match:			11.20%	Indels:	51
DB:			21	Gaps:	7

US-09-936-737A-1 (1-375) x AAG58348 (1-331)

QY	10	TTCGTGATTCTCCTCGTTCCTGGCCATGCACACTCACTACTCTTTCCAGAAAGA	69
		: : :	
Db	37	PheValAsnValPhePheCysLeu-----PheIleuLysThrArgSerSerSerGln	54
		: : :	
QY	70	CGTGAA-----GATTGTGGACGTTTACGCCAGACAGAAAATAACAGACTTCGAT	120
		: : : : : : :	
Db	55	MetGluSerIleuLeuGlucylcysPheAlaIleThrAspLysIscysHisProkspCysLeu	74
		: : : : : : :	
QY	121	AAATCTTTTAAGACGCTCCGATCTTGACGATGCATA-----	159
		: : : : : : :	
Db	75	LysAlaAsnAsnGluGlnLysPyrThrSphalacysGlnSerAlaIalaIeuValAlaVal	94
		: : : : : : :	
QY	160	-----AAAACATGTTTCAAG-----ACGAGTACTGC	186
		: : : : : : : :	
Db	95	SerIleuIleSerSerAlaArgValIlePhePylsIleaspSerLysTyrrHirGluTyrSer	114
		: : : : : : : :	
QY	187	TACATCGTTTTGAAGCACACGGTCACACAGAA-----	219
		: : : : : : :	
Db	115	ProGlnTyrIleuValAspAsnValGlyLysGluGluValGluGlyGluMetaspGlnPro	134
		: : : : : : :	
QY	220	---TGTTACTACAATGTC-----	234
		: : : : : : :	
Db	135	SerCysGlnTyrThrValGlyAsnLeuLeuSerTyrIleuValGluAsnValITrPrThLys	154
		: : : : : : :	
QY	235	-----GTTGATGGTGAAGGTTAGACCAGAAAAA-----TTTGTTCGACGAAAAC	282
		: : : : : : :	
Db	155	LysGluValArgGlnArgGluMetaspGlnArgGluPheThrValLysASPcys	174
		: : : : : : :	
QY	283	TTC 285	
Db	175	Phe 175	

RESULT 13	
AAG58347	
ID	AAG58347 standard; Protein: 336 AA.
XX	
AC	AAG58347;
XX	
DI	18-OCT-2000 (first entry)
XX	
DE	Arabidopsis thaliana protein fragment SPQ ID NO: 75508.

XX Protein identification: signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.

PM EPI033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134418.
PR 14-MAY-1999; 99US-0134421.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134768.
PR 20-MAY-1999; 99US-0134941.
PR 21-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137502.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.

PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142380.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 27-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148684.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149330.
PR 17-AUG-1999; 99US-0149338.
PR 18-AUG-1999; 99US-0149415.
PR 20-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 25-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.

```

PR 31-AUG-1999: 99US-0151438.
PR 01-SEP-1999: 99US-0151930.
PR 07-SEP-1999: 99US-0152363.
PR 10-SEP-1999: 99US-0153070.
PR 13-SEP-1999: 99US-0153758.
PR 15-SEP-1999: 99US-0154018.
PR 16-SEP-1999: 99US-0154039.
PR 20-SEP-1999: 99US-0154779.
PR 22-SEP-1999: 99US-0155139.
PR 23-SEP-1999: 99US-0155486.
PR 24-SEP-1999: 99US-0155659.
PR 28-SEP-1999: 99US-0156458.
PR 29-SEP-1999: 99US-0156596.
PR 04-OCT-1999: 99US-0157117.
PR 05-OCT-1999: 99US-0157753.
PR 06-OCT-1999: 99US-0157865.
PR 07-OCT-1999: 99US-0158029.
PR 08-OCT-1999: 99US-0158232.
PR 12-OCT-1999: 99US-0158369.
PR 13-OCT-1999: 99US-0159293.
PR 13-OCT-1999: 99US-0159294.
PR 13-OCT-1999: 99US-0159295.
PR 14-OCT-1999: 99US-0159329.
PR 14-OCT-1999: 99US-0159330.
PR 14-OCT-1999: 99US-0159331.
PR 14-OCT-1999: 99US-0159637.
PR 14-OCT-1999: 99US-0159638.
PR 18-OCT-1999: 99US-0159584.
PR 21-OCT-1999: 99US-0160741.
PR 21-OCT-1999: 99US-0160767.
PR 21-OCT-1999: 99US-0160768.
PR 21-OCT-1999: 99US-0160770.
PR 21-OCT-1999: 99US-0160814.
PR 21-OCT-1999: 99US-0160815.
PR 22-OCT-1999: 99US-0160980.
PR 22-OCT-1999: 99US-0160981.
PR 22-OCT-1999: 99US-0160989.
PR 22-OCT-1999: 99US-0161044.
PR 25-OCT-1999: 99US-0161406.
PR 25-OCT-1999: 99US-0161405.
PR 25-OCT-1999: 99US-0161406.
PR 26-OCT-1999: 99US-0161359.
PR 26-OCT-1999: 99US-0161360.
PR 26-OCT-1999: 99US-0161361.
PR 26-OCT-1999: 99US-0161920.
PR 28-OCT-1999: 99US-0161920.
PR 28-OCT-1999: 99US-0161922.
PR 28-OCT-1999: 99US-0161993.
PR 29-OCT-1999: 99US-0162142.

```

```

Alignment Scores:
Pred. No.: 6.61 Length: 336
Score: 75.50 Matches: 34
Percent Similarity: 34.75% Conservative: 15
Best Local Similarity: 24.11% Mismatches: 41
Query Match: 11.20% Indels: 51
DB: 21 Gaps: 7

```

```

US-09-936-737a-1 (1-375) x AAG58347 (1-336)
QY 10 TTCTGATTTCCTTCCTTCCTCGCAAGCTTGATCTGACTGCTTCTTCAGAAAGAA 69
DB 42 PheValAsnValPhePheCysLeu-----PheLeuLeuLysTrpArgSerSerGln 59
QY 70 CGTGAA-----GATTGTGGACGTTTACGCGAACAAGAAATATACGACTTGAT 120
DB 60 MetGluSerLeuLeuGlnCysPheAlaIleThrAspGlyLysCysHisProAspCysLeu 79
QY 121 AAATCTTTAAGAACTCTGATCTTCGACGATCAAA----- 159
DB 80 LysAlaAsnAsnGlnGlnLysPyrAspAlaCysGlnSerAlaAlaLeuValAlaVal 99
QY 160 -----AAACATGTTTCAAG-----ACGAGTACTGC 186
DB 100 SerLeuIleSerSerAlaValIlePhePheLysIleAspSerLysTyrThrGluTyrSer 119

```

```

QY 187 TACATGCTTTTGACAGACACGCTCAACAAGAA----- 219
DB 120 ProGlnTyrLeuValAsnValGlyGluGluValGluGluMetAspGlnPro 139
QY 220 ---TGTACTACATGTC----- 234
DB 140 SerCysGlnTyrThrValGlnLysLeuLeuSerTyrLeuValGluAsnValTrpThrLys 159
QY 235 -----GTTGATGCGTACAGATTAGCACAAGAAAA-----TTTGTTGTCGACGAAAC 282
DB 160 LysGluValArgGlnArgGlnMetAspGlnGlnArgGluPheThrValLysAspCys 179
QY 283 TTC 285
DB 180 Phe 180

RESULT 14
AAB14260
ID AAB14260 standard; Protein; 418 AA.
XX AC AAB14260;
XX 09-FEB-2001 (first entry)
XX DE Mouse Ice-4 protein sequence #2.
XX KW ced-3; virally induced cell death; apoptosis; gene therapy; neural;
XX KW muscular degenerative disease; myocardial infarction; stroke; aging;
XX KW interleukin-1beta converting enzyme; ICE; cysteine protease; mouse.
XX OS Mus sp.
XX PN US6083735-A.
XX PD 04-JUL-2000.
XX 10-JUN-1994; 94US-0258287.
XX PF 24-JUN-1993; 93US-0080850.
XX PR (GEHO) GEN HOSPITAL CORP.
XX PA
XX PI Yuan J, Miura M;
XX DR WPI: 2000-464343/40.
XX PT New human Ich-1L and Ich-1S proteins for negative and positive
XX PT regulation of programmed cell death and for developing therapeutic
XX PT methods for diseases and conditions characterized by cell death, e.g.
XX PT myocardial infarction or stroke
XX PS Disclosure: Fig 17; 121pp; English.
XX CC The present sequence is a mouse Ice-4 protein sequence. The present
XX CC sequence was used in a sequence homology comparison with the protein
XX CC sequences of human Ice-ced 3 homolog (Ich-1) (AAB14253), murine ICE2
XX CC interleukin-1beta converting enzyme (mICE) (AAB14249), murine ICE2
XX CC (AAB14252) and C. elegans ced-3 (AAB14246). The coding sequence of the
XX CC present sequence is a member of a family of genes involved in programmed
XX CC cell death (apoptosis). Ich-1 may play an important role in both the
XX CC positive and negative regulation of apoptosis. The Ich gene may be used
XX CC in gene therapy in disorders characterized by cell death e.g. neural and
XX CC muscular degenerative diseases, myocardial infarction, stroke, virally
XX CC induced cell death and aging.
XX SQ Sequence 418 AA;

Alignment Scores:
Pred. No.: 7.03 Length: 418
Score: 75.50 Matches: 26
Percent Similarity: 54.79% Conservative: 14
Best Local Similarity: 35.62% Mismatches: 14

```

Query Match: 11.20% Indels: 19
 DB: 21 Gaps: 6
 US-09-936-737a-1 (1-375) x AAB14260 (1-418)
 QY 193 GTTTTGAAGACACGCGTCACAAAGAAATGTTACTACATGTCGTTGATGGAAGAGTTA 252
 |||||:||||| ||| ||| |||||:|||||:|||||:|||||
 Db 26 ValPheAspAspLeuValGluLys-----AsnValLeuAsnGlyAspGluLeu 41
 |||||:|||||:|||||:|||||:|||||:|||||:|||||
 QY 253 -----GACCAAGAAAAATTGTTGTCGAC-----GAAAACCTTCACGGAATAAT 294
 |||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 42 LeuLysIleGlyGluSerAlaSerPheIleLeuAsnLysAlaGluAsnLeuValGluAsn 61
 |||||:|||||:|||||:|||||:|||||:|||||:|||||
 QY 295 TATTTG-----ACAGACTGCGAGGTAATA---GATCAGGTAATGCGGACGATACAGT 345
 |||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 62 PheLeuGluLysThrAspMetAlaGlyLysIlePheAlaGlyHisIleAlaAsnSerGln 81
 |||||:|||||:|||||:|||||:|||||:|||||:|||||
 QY 346 GACGAG-----TCAGATGAAGTTGAT 366
 |||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 82 GluGlnLeuSerLeuGlnPheSerAsnAspGluAsp 94
 |||||:|||||:|||||:|||||:|||||:|||||:|||||
 RESULT 15
 AAY50109 standard; Protein; 419 AA.
 AC AAY50109;
 XX 21-JAN-2000 (first entry)
 DE Murine caspase-12 splice variant, caspase-12L.
 XX
 KW Caspase; splice variant; truncated; programmed cell death; apoptosis;
 KW regulation; proteolytic cascade; malignant condition; cancer;
 KW premalignant condition; solid tumour; lymphoma;
 KW chronic lymphocytic leukaemia; prostatic hypertrophy;
 KW preneoplastic liver focus; chemotherapy resistance; autoimmune disease;
 ds.
 OS Mus sp.
 XX
 FH Key Location/Qualifiers
 FT Cleavage-site 94..95
 FT Protein 95..419
 FT /note= "Caspase-12"
 FT Cleavage-site 318..319
 XX
 PN W09952925-A1.
 XX
 PD 21-OCT-1999.
 XX
 PF 14-APR-1999; 99WO-US08064.
 XX
 PR 16-APR-1998; 98US-0081962.
 XX
 PA (GEHO) GEN HOSPITAL CORP.
 XX
 PI Yuan J, Morishima N;
 XX
 DR WPI: 1999-620369/53.
 DR N-PSDB: AA32645.
 XX
 PT New nucleic acid encoding the short form of caspase-12, used e.g. for
 PT treating tumors -
 XX
 PS Claim 1; Fig 1; 68pp; English.
 XX
 CC This sequence represents a cDNA encoding murine caspase-12L, a splice
 CC variant of caspase-12. Caspases are a family of proteins involved in
 CC the regulation of apoptosis and are synthesised as proforms which are
 CC activated via cleavage after specific Asp residues. Mammalian cells
 CC express several caspases, and it is thought that these act in a
 CC proteolytic cascade to cause programmed cell death. Nucleic acids
 CC encoding caspase-12S (AA32644) or truncated forms of caspase-12L

CC (AA32646, AA32647) are used for production, recombinantly or in vivo,
 CC of caspase-12 polypeptides which induce programmed cell death. This is
 CC particularly useful for treating (pre)malignant conditions (e.g., solid
 CC tumours, B cell lymphoma, chronic lymphocytic leukaemia, prostatic
 CC hypertrypthy, preneoplastic liver foci and resistance to chemotherapy), or
 CC autoimmune diseases. The caspase-12 proteins can also be used to raise
 CC specific antibodies (for example, to determine gene expression and to
 CC screen expression libraries) or as molecular weight markers. Fragments
 CC of caspase-12-encoding nucleic acids can be used as probes to isolate
 CC the caspase-12 gene (and its allelic variants); in fluorescent in situ
 CC hybridisation for chromosomal location of the caspase-12 gene, and for
 CC Northern blotting to determine caspase-12 mRNA expression in tissues.
 SQ Sequence 419 AA;
 Alignment Scores:
 Pred. NO.: 7.04 Length: 419
 Score: 75.50 Matches: 26
 Percent Similarity: 54.79% Conservative: 14
 Best Local Similarity: 35.62% Mismatches: 14
 Query Match: 11.20% Indels: 19
 DB: 20 Gaps: 6
 US-09-936-737a-1 (1-375) x AAY50109 (1-419)
 QY 193 GTTTTGAAGACACGCGTCACAAAGAAATGTTACTACATGTCGTTGATGGAAGAGTTA 252
 |||||:||||| ||| ||| |||||:|||||:|||||:|||||
 Db 26 ValPheAspAspLeuValGluLys-----AsnValLeuAsnGlyAspGluLeu 41
 |||||:|||||:|||||:|||||:|||||:|||||:|||||
 QY 253 -----GACCAAGAAAAATTGTTGTCGAC-----GAAAACCTTCACGGAATAAT 294
 |||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 42 LeuLysIleGlyGluSerAlaSerPheIleLeuAsnLysAlaGluAsnLeuValGluAsn 61
 |||||:|||||:|||||:|||||:|||||:|||||:|||||
 QY 295 TATTTG-----ACAGACTGCGAGGTAATA---GATCAGGTAATGCGGACGATACAGT 345
 |||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 62 PheLeuGluLysThrAspMetAlaGlyLysIlePheAlaGlyHisIleAlaAsnSerGln 81
 |||||:|||||:|||||:|||||:|||||:|||||:|||||
 QY 346 GACGAG-----TCAGATGAAGTTGAT 366
 |||||:|||||:|||||:|||||:|||||:|||||:|||||
 Db 82 GluGlnLeuSerLeuGlnPheSerAsnAspGluAsp 94
 |||||:|||||:|||||:|||||:|||||:|||||:|||||

Search completed: February 25, 2003, 02:25:51
 Job time : 41 secs

* * *

Wed Feb 26 15:58:50 2003

us-09-936-737a-1.ra1

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 25, 2003, 02:24:25 ; Search time 13 Seconds
(without alignments)
1697.476 Million cell updates/sec

Title: US-09-936-737A-1
Perfect score: 674
Sequence: 1 atgaagattcttctgtatc.....atgaagttgataagattaa 375

Scoring table: BLOSUM62
Xgapop 10.0, Ygapext 0.5
Xgapop 10.0, Ygapext 0.5
Xgapop 6.0, Ygapext 7.0
Delop 6.0, Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+n2p model -DEV=xlh
-O=/cgn2_1/USPRO.spool/US0936737/runcat.14022003.100558.1424/app.query.fasta.1.519
-DB=Issued Patents.AA -OPMT=fastan -SUFFIX=ra1 -MINMATCH=0.1 -LOOPEL=0
-LIST=45 -DOCALLIGN=200 -THR.SCORE=pcpt -THR.MAX=100 -THR.MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=prto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0936737 -CGN1.1.1.4 -errnat.14022003.100558.1424 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MMAP -LARGEQUERY -NEG.SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-MARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -YGAPEXT=0.5 -FGAPOP=6 -FGAEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Issued Patents.AA:*

- 1: /cgn2_6/ptodata/1/aa/5A.COMB.pep:*
- 2: /cgn2_6/ptodata/1/aa/5B.COMB.pep:*
- 3: /cgn2_6/ptodata/1/aa/6A.COMB.pep:*
- 4: /cgn2_6/ptodata/1/aa/6B.COMB.pep:*
- 5: /cgn2_6/ptodata/1/aa/PCTUS.COMB.pep:*
- 6: /cgn2_6/ptodata/1/aa/backfilest.pep:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	88.5	13.1	147	1	US-08-044-547-1
2	82.5	12.2	412	2	US-08-741-134-2
3	80.5	11.9	126	1	US-08-044-547-3
4	79.5	11.8	150	4	US-08-334-179A-10
5	79.5	11.8	1038	4	US-08-334-179A-8
6	79.5	11.2	172	4	US-09-139-600-11
7	75.5	11.2	172	4	US-09-187-789-16
8	75.5	11.2	418	3	US-08-258-287B-58
9	75.5	11.2	418	3	US-08-368-704C-56
10	75.5	11.2	419	3	US-08-258-287B-57
11	75.5	11.2	419	3	US-08-368-704C-55
12	74	11.4	1481	2	US-08-616-844-40

13	74	11.4	1481	2	US-08-599-654-40
14	74	11.4	1481	3	US-08-944-868A-40
15	74	11.4	1481	3	US-08-944-868B-40
16	74	11.4	1481	3	US-08-944-496-40
17	73.5	10.9	562	6	5258502-2
18	72.5	10.8	976	4	US-09-302-812-4
19	72.5	10.8	976	4	US-09-511-477-4
20	72.5	10.8	976	4	US-09-511-507-4
21	72.5	10.7	2555	3	US-09-058-489-36
22	71.5	10.6	146	2	US-08-888-497-34
23	71.5	10.6	146	2	US-08-362-230-41
24	71.5	10.6	146	5	PCT-US94-07926-34
25	71.1	10.5	129	4	US-09-134-001C-4545
26	71.1	10.5	450	4	US-09-254-667-1
27	70.5	10.5	849	4	US-09-157-257-4
28	69.5	10.3	124	1	US-08-170-360-1
29	69.5	10.3	450	2	US-08-170-360-1
30	69.5	10.3	450	4	US-09-265-630-11
31	69.5	10.3	355	4	US-08-961-088-200
32	68.5	10.5	355	4	US-09-403-768-6
33	68.5	10.2	425	1	US-08-713-636-2
34	68.5	10.1	425	1	US-08-700-749A-5
35	68	10.1	425	3	US-09-020-684-5
36	68	10.1	425	3	US-09-020-685-5
37	68	10.1	425	3	US-09-020-685-5
38	68	10.1	425	4	US-09-265-630-11
39	67.5	10.0	388	4	US-09-265-630-6
40	67.5	10.0	500	4	US-09-265-630-13
41	67.5	10.0	500	4	US-08-170-360-2
42	67	9.9	124	2	US-08-888-497-41
43	67	9.9	124	4	US-09-362-230-41
44	67	9.9	124	4	US-09-362-230-41
45	67	9.9	124	5	PCT-US94-07926-41

ALIGNMENTS

RESULT 1
US-08-044-547-1
; Sequence 1, Application US/08044547
; Patent No. 5324715
; GENERAL INFORMATION:
; APPLICANT: Connolly, Thomas M.
; TITLE OF INVENTION: Protein for Inhibiting
; TITLE OF INVENTION: Collagen-Stimulated Platelet Aggregation
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESS: Merck & Co., Inc.
; STREET: P.O. Box 2000
; CITY: Rahway
; STATE: New Jersey
; COUNTRY: US
; ZIP: 07065
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/08/044, 547
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA: US 07/594, 917
APPLICATION NUMBER: 09-OCT-1990
FILING DATE: 09-OCT-1990
ATTORNEY/AGENT INFORMATION:
NAME: Parr, Richard S.
REGISTRATION NUMBER: 32,586
REFERENCE/DOCKET NUMBER: 18053
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 594-4958
TELEFAX: (908) 594-4720

Sequence 40, Appl
Sequence 40, Appl
Sequence 40, Appl
Sequence 40, Appl
Patent No. 5258502
Sequence 4, Appl1
Sequence 4, Appl1
Sequence 4, Appl1
Sequence 36, Appl
Sequence 34, Appl
Sequence 34, Appl
Sequence 4, Appl1
Sequence 4, Appl
Sequence 1, Appl1
Sequence 1, Appl1
Sequence 1, Appl1
Sequence 2, Appl1
Sequence 5, Appl1
Sequence 5, Appl1
Sequence 5, Appl1
Sequence 5, Appl1
Sequence 11, Appl
Sequence 6, Appl1
Sequence 13, Appl
Sequence 2, Appl
Sequence 41, Appl
Sequence 41, Appl

```

      rec. NO.:      0.0122
      Score:         89.50
      Percent Similarity: 35.25%
      Best Local Similarity: 24.59%
      Query Match:    13.13%
      DB:             1

```

(1-147) 06-06-044-34/-1

```

Oy      1  ATGAGATATTTCGATTCCTTCCTTCCTGCTCCGACGATGCTGATCTCA----- 51
Db      1  MetAsnSerPheLeuPheSerLeuAlaCys-----SerLeuLeuAlaIleProAla 18
Oy      51  ----- 51
Db      19  IleArgAlaGlnAspGlnAspAlaGlyGlyAlaGlyAspGlnThrSerGlnGlyGlnAsp 38
Oy      52  ACTACTTCCTTCGACAGAA----- 69
Db      39  ThrThrGlySerAspGlnThrThrProSerThrGlyGlyGlyAspGlyGlyAsnGln 58
Oy      70  -----CGTAGAATTCGTGGACGTTTACGCCGACAGCAAAATATACAGAC 114
Db      59  ThrIleThrAlaGlyAsnGlnLysAspCysTrpSerIleYsArpProGlyTrpLysLeuProAsp 78
Oy      115  TTCGATTAATCTTTAAGAAGCTCTCGATCTTCAGCAATGCAGAAATAAACAATGTTTCAG 174
Db      79  AsnLeuLeuThrLysThrGlnPheThrSerValAspGlnCysArgLysMetCysGlnGln 98
Oy      175  ACGGACTAC-----TGCATCATGGTTTGAAGACACGGTCACACAGCAATGTTTAC 225
Db      99  SerAlaValaGlnProSerCysTyrIleLeuGlnIleAsnThrGlnThrAsnGlnCysTyr 118
Oy      226  TACAAT 231
Db      119  ArgAsn 120

```

RESULT 2
US-08-741-134-2

Sequence 2, Application US/08741134
Patent No. 5861498

GENERAL INFORMATION:

APPLICANT: Alnemyl, Eamad S.

APPLICANT: Fernandes-Alnemri, Teresa
TITLE OF INVENTION: IMMUNODIFFERENTIATION

TITLE OF INVENTION: AND COMPOSITIONS FOR MAKING
LEUCOPHILIN FRBP46 AND

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock

STREET: One Liberty Place - 46th floor
CITY: New York

City: Philadelphia
STATE: Pennsylvania

COUNTRY: USA
ZIP: 19103

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compati

OPERATING SYSTEM: Windows 3.11
SOFTWARE: Microsoft Word 6.0

CURRENT APPLICATION DATA: for windows 6.1

```

1 APPLICATION NUMBER: US/08/741,134
2 FILING DATE:
3 CLASSIFICATION: 424
4 PRIOR APPLICATION DATA:
5 APPLICATION NUMBER: US 60/007,163
6 FILING DATE: 01-NOV-1995
7 CLASSIFICATION: 424
8 ATTORNEY/AGENT INFORMATION:
9 NAME: DeLuca, Mark
10 REGISTRATION NUMBER: 33,229
11 REFERENCE/DOCKET NUMBER: TJU-2090
12 TELECOMMUNICATION INFORMATION:
13 TELEPHONE: 215-568-3100
14 TELEFAX: 215-568-3439
15 INFORMATION FOR SEQ ID NO:
16 SEQUENCE CHARACTERISTICS:
17 LENGTH: 412 amino acids
18 TYPE: amino acid
19 TOPOLOGY: linear
20 MOLECULE TYPE: protein
21 US-08-741-134-2

```

Alignment Scores:
Pred. No.

Score:	0.0899
Percent Similarity:	82.50
Best Local Similarity:	41.74%
Query Match:	27.83%
DB:	12.24%
	2

US-08-741-134-2 (1-412)

```

QY 43 CTGATTCACACTACTCTCTTCGAGAAAGCAAGCTGAAGATTGTGGACGTTTACGCGAAACGA 102
Db 116 LeuValProAlaIAsnLysAsnLysArgLysLeuGlnAsnAlaAsnAspAlaThrIAsnLys 139
QY 103 AAATATACAGACTTCGATTAATCTTTTAAAGATCC-----TCGATCTT 147
Db 136 LysAlaLysProAspLysLysAlaGlyLysAsnSerAlaProAlaIAsnLysSer 155
QY 148 GACGAAATCAAAAAAACATGTTTCAAGACGAGATCTACATCGTTTGAAGACAG 207
Db 156 AspAspAspAsp-----GlnAspGln 162
QY 208 GTCAACACAGGAATGTTACTACAATGCTGTGATGCTGAAGACTTGACACCAAGAAAAATTT 267
Db 163 LeuGlnLys-----PheLeuAspLysGlnAspLysAspThrAspGln----- 176
QY 268 GTTGTGCGAAGAAACACTTCGCGAAAAATATTATTCAGAGCTGCGAGCGTAAAGATGCGCGT 327
Db 177 ---AsnAspLysSerPheLysMetAsn-----ThrSerAlaGlnGlyLysAspAspSerAsp 193
QY 328 AATGCGGACGATACAGGTGACGAGCTACGATGACAGTGTGATGAGAAAT 372
Db 194 GlnGlnAspAspAspGlnAspGlnGlnGlnAspGlnGlnGlnAspAspAsp 208

```

US-08-044-547-3

; Sequence 3, Application US/08044547
; Patent No. 5324715

GENERAL INFORMATION:

APPLICANT: Connolly, Thomas M
APPLICANT: Keller, Paul M

TITLE OF INVENTION: Protein for Inhibiting

NUMBER OF SEQUENCES:	12
Collagen-Stimulated Platelet Aggregation	

CORRESPONDENCE ADDRESS:
ADDRESSEE: Merck & Co Inc

STREET: P.O. Box 2000
CITY: Del...

STATE: New Jersey

COUNTRY: US
ZIP: 07065

```

1 APPLICATION NUMBER: US/08/334,179A
2 FILING DATE: 04-NOV-1994
3 CLASSIFICATION: 435
4 ATTORNEY/AGENT INFORMATION:
5 NAME: CORSTANJE, BRAHM J.
6 REGISTRATION NUMBER: 34,804
7 REFERENCE/DOCKET NUMBER: 5473
8 TELECOMMUNICATION INFORMATION:
9 TELEPHONE: 513-627-2858
10 TELEFAX: 513-627-0260
11 INFORMATION FOR SEQ ID NO: 10:
12 SEQUENCE CHARACTERISTICS:
13 LENGTH: 150 amino acids
14 TYPE: amino acid
15 TOPOLOGY: linear
16 MOLECULE TYPE: protein
17 US-08-334-179A-10
18
19 Alignment Scores:
20 Pred. No.: 0.152 Length: 150
21 Score: 79.50 Matches: 27
22 Percent Similarity: 40.46% Conservative: 26
23 Best Local Similarity: 20.61% Mismatches: 35
24 Query Match: 11.80% Indels: 43
25 Gaps: 4
26 DB:
27
28 US-09-936-737A-1 (1-375) x US-08-334-179A-10 (1-150)
29
30 QY 10 TTCCTGATTTCCTTCCTTGCCCTGGCAAGCTGGTGATTCACACTACTCTTCAGAGAGA 69
31 ||| :::: ||||| ||| |||||:::|||||::: |||
32 9 PheAGValaProTyrPLeuTPrLalValLeuValSerThrhrLalalaSerGln 28
33
34 QY 70 CGGAAAGATTGTGGAGCGTTTACGCGAACAAGAAATATATACAGACTTCGATAAATT 129
35 ||| :::: ||||| :::: |||
36 29 AsnGlnGlu-----ArgLeuGlyAlaPheLysAspProTyr 40
37
38 QY 130 AAGAGCTCTTGACTTGACGAATGCAAAAAGCA-----TGT 168
39 :::: ||| :::: |||
40 41 GlnGlnAspLeuGlyLeuGlyIleuSerArgIleSerHisGlnAsnGlyThrIleLeuGly 60
41
42 QY 169 TTCAGACGAGAGTACGCTACATCGTTTGA-----GACACGGTC 210
43 ||| ||||| :::: |||
44 61 SerLysGlySerThrCysTyrGlyLeuTyrGlnLysSerLysGlyAspIleAsnLeuVal 80
45
46 QY 211 AACAGCAAGTTACTACCAATGCGTGTGATGGTGAAGAGTTAGACCAAGAAATTTGTT 270
47 :: ||||::: ||| :::: |||
48 81 LysGlnIleCysTyrSerHisIleGlyAspProGlnGlnCysHisTyrGlnGlnCysVal 100
49
50 QY 271 GTC----- 273
51 |||
52 101 ValIhrThrThrProProSerIleGlnAsnGlyThrTyrArgPheCysCysCysSerThr 120
53
54 Db 274 -----GACGAAACTTCACGAAATATAT 297
55 :: |||||:::|||||:::
56 121 AspLeuGlyAsnValaAsnPheThrGlnAsnPhe 131
57
58 RESULT 5
59 US-08-334-179A-8
60 Sequence 8, Application US/08334179A
61 Patent No. 6306622
62 GENERAL INFORMATION:
63 APPLICANT: ROSENBAUM, JAN S.
64 APPLICANT: NOHNO, TSUTOMI
65 TITLE OF INVENTION: cDNA ENCODING A BMP TYPE II RECEPTOR
66 NUMBER OF SEQUENCES: 14
67 CORRESPONDENCE ADDRESSES:
68 ADDRESSEE: THE PROCTOR AND GAMBLE COMPANY
69 STREET: 11810 EAST MIAMI RIVER ROAD
70 CITY: OHSS
71 STATE: OH
72 COUNTRY: US
73 ZIP: 45061
74 COMPUTER READABLE FORM:

```

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.30, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/334,179A
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: CONSTANJE, BRAHM J.
REGISTRATION NUMBER: 34,804
TELECOMMUNICATION INFORMATION:
TELEPHONE: 513-627-2858
TELEFAX: 513-627-0260
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1038 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-334-179A-8

Alignment Scores:
Pred. No.: 0.276
Score: 79.50
Percent Similarity: 40.46%
Best Local Similarity: 20.61%
Query Match: 11.80%
DB: 4 Gaps: 4

US-09-936-737A-1 (1-375) x US-08-334-179A-8 (1-1038)
QY 10 TTCTTGATTCCTTCCTTCCTGCGCAAGCTTGCTGATCTCACTACTTTCAGAGAA 69
DB 9 PheAglValProTPrLeuTPrAlaValLeuValSerThrThrAlaAlaSerGln 28
QY 70 CGTGAAGATGTTGGACGCTTTTACGACAGCAAGAAATATACAGACTTCGATAATCTTTT 129
DB 29 AsnGlnGlu-----ArgLeuCysAlaPheLysAspProTyr 40
QY 130 AAGAAGTCCTGATCTTCGCAATGCAAAAAACA-----TGT 168
DB 41 GlnGlnAspLeuGlyIleGlyGlnSerArgIleSerHisGlnAsnGlyThrIleLeuGly 60
QY 169 TTCAGACGAGTACTGCTCATCGTCTTTTGA-----GACACGGTC 210
DB 61 SerLysGlySerThrCysTyrGlyLeuTPrGlyLysSerLysGlyAspIleAsnLeuVal 80
QY 211 AACAGAGATGTTACTACAAATGCTGTTGATGGTGAAGATTAGACCAAGAAATTTGTT 270
DB 81 LysGlnIleLysTPrSerHisIleGlyAspProGlnGlnCysHisTyrGlnGlnCysVal 100
QY 271 GTC----- 273
DB 101 ValIThrThrProProSerIleGlnAsnGlyThrTyrArgPheCysCysSerThr 120
QY 274 -----GACGAAACTTCACGGAATAATTAT 297
DB 121 AspLeuCysAsnValAsnPheThrGlnAsnPhe 131

RESULT 6
US-09-187-789-16
Sequence 16, Application US/09187789
Patent No. 6340740
GENERAL INFORMATION:
APPLICANT: Alnemri, Emdad S.
APPLICANT: Fernandez-Alnemri, Teresa
TITLE OF INVENTION: CASPASE-14, AN APOPTOTIC PROTEASE, NUCLEIC ACID ENCODING
FILE REFERENCE: 480140.434C1
CURRENT APPLICATION NUMBER: US/09/187,789
CURRENT FILING DATE: 1998-11-06

```

```

NUMBER OF SEQ ID NOS: 78
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 16
LENGTH: 172
TYPE: PRT
ORGANISM: Mus musculus
US-09-187-789-16

Alignment Scores:
Pred. No.: 0.487
Score: 75.50
Percent Similarity: 54.79%
Best Local Similarity: 35.62%
Query Match: 11.20%
DB: 4 Gaps: 6

US-09-936-737A-1 (1-375) x US-09-187-789-16 (1-172)
QY 193 GTTTTGAAGACAGCGTCAACAGAAATGTTACTACAAATGCTGTTGATGGTGAAGATT 252
DB 26 ValPheAspAspLeuValGlnLys-----AsnValLeuAsnGlyAspGlnLeu 41
QY 253 -----GACCAAGAAAAATTTGTTGCGAC-----GAAACTTCACGGAAT 294
DB 42 LeuLysIleGlyGlnSerAlaSerPheIleLeuAsnLysAlaGlnAsnLeuValGlnAsn 61
QY 295 TATTTG-----ACAGACTGCGAGGCTAAA--GATGCGAGGTTAATGCGGACGTAAGGT 345
DB 62 PheLeuGlnLysThrAspMetAlaGlyLysIlePheAlaGlyHisIleAlaAsnSerGln 81
QY 346 GAGCAG-----TCAGATGAAGTTGAT 366
DB 82 GlnGlnLeuSerLeuGlnPheSerAsnAspGlnAsp 94

RESULT 7
US-09-139-600-11
Sequence 11, Application US/09139600
Patent No. 6432628
GENERAL INFORMATION:
APPLICANT: Alnemri, Emdad S.
APPLICANT: Fernandez-Alnemri, Teresa
TITLE OF INVENTION: CASPASE-14, AN APOPTOTIC PROTEASE, NUCLEIC ACID ENCODING
FILE REFERENCE: 480140.434
CURRENT APPLICATION NUMBER: US/09/139,600
NUMBER OF SEQ ID NOS: 65
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 11
LENGTH: 172
TYPE: PRT
ORGANISM: Mus musculus
US-09-139-600-11

Alignment Scores:
Pred. No.: 0.487
Score: 75.50
Percent Similarity: 54.79%
Best Local Similarity: 35.62%
Query Match: 11.20%
DB: 4 Gaps: 6

US-09-936-737A-1 (1-375) x US-09-139-600-11 (1-172)
QY 193 GTTTTGAAGACAGCGTCAACAGAAATGTTACTACAAATGCTGTTGATGGTGAAGATT 252
DB 26 ValPheAspAspLeuValGlnLys-----AsnValLeuAsnGlyAspGlnLeu 41
QY 253 -----GACCAAGAAAAATTTGTTGCGAC-----GAAACTTCACGGAAT 294
DB 42 LeuLysIleGlyGlnSerAlaSerPheIleLeuAsnLysAlaGlnAsnLeuValGlnAsn 61
QY 295 TATTTG-----ACAGACTGCGAGGCTAAA--GATGCGAGGTTAATGCGGACGTAAGGT 345
DB 295 TATTTG-----ACAGACTGCGAGGCTAAA--GATGCGAGGTTAATGCGGACGTAAGGT 345

```


[illegible]

us-09-936-737a-1.rai

Page 8

1.44	Length:	1481
74.00	Mass:	

```

Length:      1481
Matches:      43
Conservative: 14
Mismatchches: 39
Indels:       50
Gaps:         7

```

```
D6      |||||::||| | ||||| ||| |||||
764 SerSerSerSerSerSerSerSerClnDnDnDnT - 32
```

323 -----GCATCTTTACCTCGCAGTCTGTCAATATTTCCGTGAAG 28

261 TTTTCGTCGACACAAATTTTCTTGCTACTCT---TCACCA-----TCAACG 234

233 ACATTGTAGTACATTCCTTGTGACCGTGTCTTCAAAAACGATGTACAGTACTCC--- 177

1/6 GTCTGAACAT-----165

164 -----GTTTTTTCATTCGTCAGATCAGAGACTTCTTAAAGATTATCG 117

116 AAGTCTATATTTCTGTCCGTAACGTCACATCTTCACGTTCTTCTGAAGAA 57

56 GTAGTTGAGATCAGCAAG 39

RESULT 14
5-08-944-8692-10

GENERAL INFORMATION:

NUMBER OF SEQUENCES:	54	TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
NUMBER OF INVENTION:		
TITLE OF INVENTION:		

STREET: 1155 Avenue of the Americas
CITY: New York

ZIP: 10036-2711
COMPUTER READABLE FORM.

OPERATING SYSTEM: PC-DOS/MS-DOS

APPLICATION NUMBER: US/08/944,868A
FILING DATE:

APPLICATION NUMBER: 08/599,654
FILING DATE:

FILED DATE: 10-FEB-1995

REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7853-041

TELEFAX: (212) 869-8864
TELEX: 66141 PENNTE

LENGTH: 1481 amino acids

MOLECULE TYPE: protein

ment Scores:

Local Similarity: 29.4

-936-737A-1 (1-375) x U

```

|||||:|:|
764 serverserver

```

784 SerHisHisLeuPheSer

```

804 SerThrSerAsn] assert
      :::| | |      :::

```

ThrThrSerThrLeuAlaLeu
111 :::
824

344 mbwDnewall T

[illegible]

16 AAGTCCTGTAATTTCTCTGT

56 GTAGTTGAGATCAGCAAG 3

15
44-423A-10

GENERAL INFORMATION:

NUMBER OF SEQUENCES: 54

STREET: 1155 Avenue of
CITY: New York

STREET: 1155 Avenue of
CITY: New York

```

Db      880      -----SerGlnSerThrProHisGlnGlnLys 888
OY      56      GTAGTGGAGATGAGCAAG 39
          |||:::      |||||
Db      889      ValIleThrGlnSerLys 894

Search completed: February 25, 2003, 02:29:51
Job time : 17 secs

```

Search completed: February 25, 2003, 02:29:51
Job time : 17 secs

...

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 25, 2003, 02:25:55 ; Search time 31 Seconds

(without alignments)
751.689 Million cell updates/sec

Title: US-09-936-737a-1

Perfect score: 674

Sequence: 1 atgaagattcttgattc.....atgaagtgaatgaattaa 375

Scoring table:
BL0SUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 156504 segs, 31069816 residues

Total number of hits satisfying chosen parameters: 313008

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_n2p_model -DEV=xlh
-O=/cgn2_1/USPTO.spool/US09936737/rnat_14022003_100558_1452/app_query.fasta_1.519
-DB=Published Applications_AA -QFMT=fastan -SUFFIX=rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=bl0sum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09936737@cgn_1_1_2_@rnat_14022003_100558_1452
-NCPU=6 -ICPU=3 -NO_XLPHY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications_AA:

1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description	
1	75.5	11.2	172	US-09-989-903-16	
2	74	11.0	1178	US-10-041-856-9	
3	74	11.4	1481	US-09-371-900-40	
c	4	74	11.4	1481	US-09-924-417-60

Result No.	Score	Query Match Length	ID	Description
5	73.5	10.9	562	US-09-801-368-82
6	72.5	10.8	976	US-08-973-451-4
7	72	10.7	143	US-09-867-550-2038
8	71.5	10.6	156	US-09-833-790-252
9	70.5	10.5	578	US-09-821-839-2
10	70	10.4	1167	US-09-815-242-11522
11	69.5	10.3	674	US-09-765-272-200
12	68.5	10.5	72	US-09-764-660-558
13	68.5	10.2	206	US-09-738-626-6537
14	68	10.1	146	US-09-993-999-7
15	68	10.4	556	US-09-801-368-426
16	67.5	10.4	281	US-09-895-913A-282
17	67.5	10.0	1036	US-09-918-508-4
18	67.5	10.0	1092	US-10-135-322-18
19	67	9.9	322	US-09-872-523-5
20	66.5	9.9	2789	US-09-801-574-57
21	66	9.8	172	US-10-087-195-6
22	66	9.8	172	US-10-085-572-6
23	66	9.8	485	US-10-140-372-11
24	65.5	9.7	258	US-09-808-602-110
25	65.5	9.7	270	US-09-816-028A-39
26	65	9.6	374	US-09-820-004-2
27	65	9.6	417	US-09-820-004-4
28	65	9.6	417	US-09-820-004-5
29	65	9.6	417	US-09-820-004-6
30	65	10.0	661	US-10-094-214-5
31	65	9.6	1125	US-09-974-296-114
32	65	10.0	2344	US-09-815-242-12713
33	65	9.6	26926	US-09-759-508B-2
34	64.5	9.6	156	US-09-925-297-511
35	64.5	9.6	977	US-09-973-451-2
36	64.5	9.6	1038	US-09-908-500A-2
37	64	9.5	225	US-10-185-770-4
38	64	9.5	272	US-10-125-852-23
39	64	9.5	292	US-10-125-852-25
40	64	9.5	467	US-09-745-763-166
41	64	9.5	292	US-09-416-344A-7
42	64	9.5	590	US-10-042-417-52
43	63.5	9.8	376	US-09-801-368-206
44	63.5	9.4	571	US-09-925-301-1031
45	63.5	9.4	829	US-09-946-805-8

ALIGNMENTS

RESULT 1
US-09-989-903-16
; Sequence 16, Application US/09989903
; Patent No. US20020146804A1
; GENERAL INFORMATION:
; APPLICANT: Alnemri, Emdad S.
; APPLICANT: Fernandez-Alnemri, Teresa
; TITLE OF INVENTION: CASPASE-14, AN APOPTOTIC PROTEASE, NUCLEIC ACID ENCODING
; FILE REFERENCE: 480140.434D1
; CURRENT APPLICATION NUMBER: US/09/989,903
; CURRENT FILING DATE: 2002-04-11
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 16
; LENGTH: 172
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-989-903-16

Alignment Scores:

Pred. No.: 0.84
Score: 75.50
Percent Similarity: 54.79%
Best Local Similarity: 35.62%
Query Match: 11.20%
DB: 10
Length: 172
Matches: 26
Conservative: 14
Mismatch: 14
Indels: 19
Gaps: 6

US-09-936-737a-1 (1-375) x US-09-989-903-16 (1-172)

OY 193 GTTTTGAAGACACGCTCAAGAAATTTACTACTGTCGTCAGGTGAAGTTA 252
 |||||:||||| ||| ||| |||||:|||||:|||||:|||||
 Db 26 ValPheAspSerLeuValGluLys-----AsnValLeuAsnGlyAspGluLeu 41
 OY 253 -----GACCAGAAAAATTTGTGTGCAC-----GAAACTTCACGAAAT 294
 |||||:||||| |||||:||||| ||||| |||||
 Db 42 LeuLysIleGlyGluSerAlaSerPheIleLeuAsnLysAlaGluAsnLeuValGluAsn 61
 OY 295 TATTTG-----ACAGACTCGCAGGTAA---GATGCAAGTAATCGCGCAGGTACAGGT 345
 :|||:||||| ||||| ||||| |||||:|||||:|||||
 Db 62 PheLeuGluLysThrAspMetAlaGlyLysIlePheAlaGlyHisIleLeuAsnSerGln 81
 OY 346 GACGAG-----TCAGATGAAGTTGAT 366
 :|||:||||| ||| ||| |||
 Db 82 GluGlnLeuSerLeuGlnPheSerAsnAspGluAspAsp 94

RESULT 2

US-10-041-856-9
 ; Sequence 9, Application US/10041856
 ; Patent No. US20020169299A1
 ; GENERAL INFORMATION:
 ; APPLICANT: SLAUGENHAUPT, SUSAN
 ; APPLICANT: GUSELA, JAMES F.
 ; TITLE OF INVENTION: GENE FOR IDENTIFYING INDIVIDUALS WITH FAMILIAL
 ; FILE REFERENCE: 1829-4004US1
 ; CURRENT APPLICATION NUMBER: US/10/041,856
 ; CURRENT FILING DATE: 2002-07-08
 ; PRIOR APPLICATION NUMBER: 60/260,080
 ; PRIOR FILING DATE: 2001-01-06
 ; NUMBER OF SEQ ID NOS: 88
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 9
 ; LENGTH: 1178
 ; TYPE: PRT
 ; ORGANISM: Caenorhabditis elegans
 ; US-10-041-856-9

Alignment Scores:
 Pred. No.: 2.14 Length: 1178
 Score: 74.00 Matches: 23
 Percent Similarity: 41.24% Conservative: 17
 Best Local Similarity: 23.71% Mismatches: 37
 Query Match: 10.98% Indels: 20
 DB: 9 Gaps: 3

US-09-936-737a-1 (1-375) x US-10-041-856-9 (1-1178)

OY 19 TCTTTCCTTTCCTGCAAGCTGCTGATCTCACTACTCTTTCAGAAAGTGA--- 75
 ||| |||||:|||||:|||||:|||||:|||||
 Db 400 SerIleGlnCysLeuSerAspIleValAlaLysThrIleSerThrIleHisValHisVal 419
 OY 76 -----GATTGTGACGTTTACGGGAACAGAAATATATACA 111
 ||| ||| |||||:||||| |||||:|||||
 Db 420 IleThrSerAspTrpLysIleIleSerCysMetLeuPhePheLysLysLysArgAsn 439
 OY 112 GACTTCGATAATCTTTTAAGAAAGTCTGATCTTCGCAAGTCAAAAACATGTTTC 171
 :|||:||||| |||||:||||| ||||| |||||
 Db 440 TyrSerAsnProPhePheArgLysLysTyrIleLeuGluIleLeuLysValProSerHis 459
 OY 172 AAGACGAGTAC---TGTACATCGTTTTCGAAACACGCTCAACAAAGAAATGTACTAC 228
 |||||:||||| ||| ||| |||||:|||||
 Db 460 LysThrTyrPheAlaCysPheAlaValSerGlnAspThr----- 472
 OY 229 AATGTCGTTGATGTAAGAAAGTTCAGCAAGAAATTTGTTCGACGAA 279
 |||||:||||| |||||:||||| |||||:|||||
 Db 473 -----AspGlyTyrLysPheAsnSerAspArgAlaSerIleAspGln 486

RESULT 3
 US-09-371-900-40
 ; Sequence 40, Application US/09371900

Patent No. US20020137700A1

GENERAL INFORMATION:

APPLICANT: FALB, DEAN A
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
 TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE

NUMBER OF SEQUENCES: 54
 CORRESPONDENCE ADDRESS:
 ADDRESSER: PENNIE & EDMONDS LLP
 STREET: 1155 Avenue of the Americas
 CITY: New York
 STATE: New York
 COUNTRY: USA
 ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/371,900
 FILING DATE: 11-Aug-1999
 CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/599,654
 FILING DATE: 09-FEB-1996
 APPLICATION NUMBER: US 08/485,573
 FILING DATE: 07-JUN-1995
 APPLICATION NUMBER: US 08/386,844
 FILING DATE: 10-FEB-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: CORUZZI, LAURA A
 REGISTRATION NUMBER: 30,742

REFERENCE/DOCKET NUMBER: 7853-104
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 790-9090
 TELEFAX: (212) 869-8864
 TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 40:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1481 amino acids
 TYPE: amino acid
 STRANDEDNESS: <Unknown>
 TOPOLOGY: unknown

MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 40:
 US-09-371-900-40

Alignment Scores:

Pred. No.: 2.27 Length: 1481
 Score: 74.00 Matches: 43
 Percent Similarity: 39.04% Conservative: 14
 Best Local Similarity: 29.45% Mismatches: 39
 Query Match: 11.37% Indels: 50
 DB: 10 Gaps: 7

US-09-936-737a-1 (1-375) x US-09-371-900-40 (1-1481)

OY 371 TCTTTCATCACTCATCTGACCTGACCTGACCTGCGCATTTACT----- 324
 |||||:|||||:|||||:|||||:|||||
 Db 764 SerSerSerSerSerSerSerSerGlyProProLeuProLeuProLeuProValSerGln 783
 OY 323 -----GCATCTTACCCCTGCGAGTCTGTCAAATAATTTTCGCTGAAG 282
 :|||:||||| |||||:||||| |||||:|||||
 Db 784 SerHisHisLeuPheSerSerIleLeuProSerThrArgAlaSerValHisLeuLeuLys 803
 OY 281 TTTTGTGCAACAATAATTTTCTGTCGCTTAACCT---TCAACA-----TCAACG 234
 :|||:||||| ||| ||| |||||:||||| |||||
 Db 804 SerThrSerAspAlaSerThrProTrpSerSerSerProSerProLeuProValSerLeu 823
 OY 233 ACATGTGATACATTCCTGTTGACCGCTGCTTCAAAAAGCATGATACAGTACTCC--- 177
 ||| |||:|||||:|||||:|||||:|||||
 Db 824 ThrThrSerThrSerAlaProLeuSerValSerGlnThrThrLeuProGlnSerSerSer 843

US-09-924-417-60	Alignment Scores:	2.27	Length:	1481
Pred. No.:	74.00	Matches:	43	
Score:	39.04%	Conservative:	14	
Percent Similarity:	29.45%	Mismatches:	39	
Best Local Similarity:	11.37%	Indels:	50	
Query Match:	10	Gaps:	7	
DB:				
US-09-936-737A-1 (1-375) x US-09-924-417-60 (1-1481)				
OY	371 TCTTCATCAAACTTATCTGACCTGTCACCTGTACCTGCCGATTACCT			324
Db	764 SerSerSerSerSerSerSerSerSerGlyProProLeuProLeuProSerValSerGln			783
OY	323 -----GCATCTTTACCCCTGCAGCTGTCAAAATTAATTTCCGCGAG			282
Db	784 SerHisHisLeuPheSerSerIleLeuProSerThrArgAlaSerValHisLeuLeuLys			803
OY	281 TTTTTCGACACAAATTTTCTGGCTTAACCTC---TCACCA-----TCACG			234
Db	804 SerThrSerSerAlaSerThrProTirSerSerSerProSerProLeuProValSerLeu			823
OY	233 ACATGTAGTACATCTTCCTGTGACCGCTGTCTCAAAAACGATTAACAGACTCC			177
Db	824 ThrThrSerThrSerAlaProLeuSerValSerGlnThrLeuProGlnSerSer			843
OY	176 -----GCTCTGAACAT-----			165
Db	844 ThrProValLeuProArgAlaArgGluThrProValThrSerPheGlnThrSerThrMet			863
OY	164 -----GTTTTTTCATTGCTGCACATCAGACGACTCTTTAAAGATTATCG			117
Db	864 ThrSerPheMetThrMetLeuHisSerSerGlnThrAlaAsp---LeuLys			879
OY	116 AAGCTGTGATATTTCCTGTTCTCGGCTAAACGTCACCAATCTTCACGCTTCTTCGAGA			57
Db	880 -----SerGlnSerThrProHisGlnGluLys			888
OY	56 GTAGTTCAGATCAGCAG 39			
Db	889 ValIleThrGluSerLys 894			
RESULT 5				
US-09-801-368-82				
Sequence 82, Application US/09801368				
Patent No. US20020128250A1				
GENERAL INFORMATION:				
APPLICANT: Busby, Robert				
APPLICANT: Call, Brian				
APPLICANT: Hecht, Peter				
APPLICANT: Holzman, Doug				
APPLICANT: Madden, Kevin				
APPLICANT: Maxon, Mary				
APPLICANT: Milne, Todd				
APPLICANT: No. US20020128250A1man, Thea				
APPLICANT: Royer, John				
APPLICANT: Salama, Sofie				
APPLICANT: Sherman, Amir				
APPLICANT: Silva, Jeff				
APPLICANT: Summers, Eric				
TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi				
FILE REFERENCE: 109272.147				
CURRENT APPLICATION NUMBER: US/09/801,368				
CURRENT FILING DATE: 2001-03-07				
PRIOR FILING DATE: 2000-01-19				
PRIOR APPLICATION NUMBER: US 09/487,558				
PRIOR FILING DATE: 1998-10-20				
NUMBER OF SEQ ID NOS: 440				
SOFTWARE: PatentIn version 3.0				
SEQ ID NO 82				

LENGTH: 562
 TYPE: PRT
 ORGANISM: Saccharomyces cerevisiae
 US-09-801-368-82

Alignment Scores:

Pred. No.: 1.99
 Score: 73.50
 Percent Similarity: 27
 Best Local Similarity: 39.81%
 Query Match: 25.00%
 DB: 10.91%
 Gaps: 1

US-09-936-737a-1 (1-375) x US-09-801-368-82 (1-562)

QY 29 GCCTCCGACGCTTGTGATCTCACTCTCTGCAAGAACGTAAGATTTGGACGT 88
 Db 311 AIsaSerGlnThrAlaThrThrValAlaThrSerLysThrSerAlaAlaSerThrSer 330
 QY 89 TTTACGCGACGAAATATACAGACTTCGATTAATCTTTAAGATCTCTGATCTTG 148
 Db 331 SerAlaSerThrSerSerAlaSerThrSerGlnLysLysThrThrGlnSerThrThrSer 350
 QY 149 AGGATTCGAAAAACATGTTTCAAGACGAGTACGTAACGTTTGAAGACACGG 208
 Db 351 ThrGlnSerLysSerLysValThrLeuSerProThrAlaSerSerAlaLleLysThrSer 370
 QY 209 TCACACGAGATGTTACTACATGCTGATGCTGAAGAGTTAACCAGAAATTTG 268
 Db 371 IleThrGlnThrThrLysThrLysThrLysThrLysThrLysThrLysThrLysThr 268
 QY 269 TTTGCGACGAAACTTCACGCGAAATTTATTTGACAGACTGCGAGGTAAGATGACGTA 328
 Db 388 LeuGlyThrThrThrThrLysThrLysThrLysThrLysThrLysThrLysThrLysThr 407
 QY 329 ATGCGGAGGTACAGGTGACGAGT 352
 Db 408 ThrLeuSerSerGlnIleThrSer 415

RESULT 6

US-09-973-451-4
 Sequence 4, Application US/09973451
 Patent No. US20020132328A1
 GENERAL INFORMATION:
 APPLICANT: JACOBSON, Myron K.
 APPLICANT: JACOBSON, Elaine L.
 APPLICANT: AM, Jean-Christophe
 APPLICANT: LIN, Winston
 TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOPHYLASE
 TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
 FILE REFERENCE: N/A
 CURRENT APPLICATION NUMBER: US/09/973,451
 PRIOR APPLICATION NUMBER: 2001-10-09
 PRIOR FILING DATE: US/09/302,812
 PRIOR APPLICATION NUMBER: 60/083,768
 PRIOR FILING DATE: 1998-05-01
 NUMBER OF SEQ ID NOS: 38
 SEQ ID NO 4
 LENGTH: 976
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-09-973-451-4

Alignment Scores:

Pred. No.: 3.04
 Score: 72.50
 Percent Similarity: 38.89%
 Best Local Similarity: 25.93%
 Query Match: 10.76%
 Length: 976
 Matches: 28
 Conservative: 14
 Mismatches: 35
 Indels: 31

DB: 10 Gaps: 5

US-09-936-737a-1 (1-375) x US-09-973-451-4 (1-976)

QY 49 TCACACTCTCTCTGCAAGAACGTAAGATTTGGACGTTT-----TAC 93
 Db 197 SerAlaSerThrSerLysThrLysThrLysThrLysThrLysThrLysThrLysThrLys 216
 QY 94 GCGACACGAAATATACAGACTTCGATTAATCTTTAAGATCTCTGATCTTGACGAA 153
 Db 217 AlaAlaLysGlnThrThrGlnLysAlaThrGlnLysThrLysThrLysThrLysThrLys 236
 QY 154 TCACACGAAATATACAGACTTCGATTAATCTTTAAGATCTCTGATCTTGACGAA 213
 Db 237 CysSerLysSerCysHisProGlnLysPys----- 247
 QY 214 AAGCAATGTTCTCAATGCTGATGTAAGATTAACCAAGAAATTTGTTGTC 273
 Db 248 AlaSerCysGlnLysPysGlnLysPys-----ValVal 258
 QY 274 GACGAAACCTTCACGCGAAATTTATTTGACAGACTCGAGGTAAGATGACGATATGCG 333
 Db 259 -----ProLysSerProLysSerPysValGlySerGlnAsp----- 270
 QY 334 GCAGTACAGGTGACGATCAGAT 357
 Db 271 ValGlyThrGlySerLysAsnAsp 278

RESULT 7

US-09-867-550-2038
 Sequence 2038, Application US/09867550
 Patent No. US20020082206A1
 GENERAL INFORMATION:
 APPLICANT: Leach, Martin D.
 APPLICANT: Mehraban, Foad
 APPLICANT: Conley, Pamela
 APPLICANT: Law, Debbie
 TITLE OF INVENTION: No. US20020082206A1 Polynucleotides from Atherogenic Cells a
 TITLE OF INVENTION: Thereby
 FILE REFERENCE: 21402-013 (Cura-313)
 CURRENT APPLICATION NUMBER: US/09/867,550
 PRIOR FILING DATE: 2001-09-20
 PRIOR APPLICATION NUMBER: USSN 60/208,427
 PRIOR FILING DATE: 2000-05-30
 NUMBER OF SEQ ID NOS: 2125
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 2038
 LENGTH: 143
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-09-867-550-2038

Alignment Scores:

Pred. No.: 2.05
 Score: 72.00
 Percent Similarity: 42.40%
 Best Local Similarity: 23.20%
 Query Match: 10.68%
 Length: 143
 Matches: 29
 Conservative: 24
 Mismatches: 60
 Indels: 12
 Gaps: 5

US-09-936-737a-1 (1-375) x US-09-867-550-2038 (1-143)

QY 10 TTTCTGATTTCTCTCTCTGCTGCAAGCTTCGATCTCACTCTTCTTGAAGAA 69
 Db 15 TTTCTGATTTCTCTCTCTGCTGCAAGCTTCGATCTCACTCTTCTTGAAGAA 69
 QY 70 CGTGAAGATTTGAGCTTTTACCGCAGACGAAATATACAGACTTCGATTAATCTTT 129
 Db 34 -----CysLeuHisThrPheCysLys-----ThrCysLleValGlnHisPhe 47
 QY 130 AAGAGTCTCTGATCTTCGACGAATTCGCAAAACATGTTTCAAGCGAGTACTGCTAC 189
 Db 130 AAGAGTCTCTGATCTTCGACGAATTCGCAAAACATGTTTCAAGCGAGTACTGCTAC 189

```

Db      48  GUAsperasnaspcysproargCysglyasnGlnValHisGluThrAsnProLeuGlu 67
QY      190  ATCGTTTAAAGACACGCGTCAACAGATGTTACTACAAATGCTGATGCT----- 243
      :      :      :      :      :      :      :      :      :      :
Db      68  MetLeuArgLeuAspsnThrLeuGluIleIlePheLysLeuValProLysLeuArg 87
QY      244  ---GAAGAGTTAGACCAAGAA---AAATTTGTTGTCGACGAAAACTTCACGAAATATAT 297
      :      :      :      :      :      :      :      :      :      :
Db      88  GUGInGluLeuGlnArgGlnSerGluPheTrpLysAsnLysLeuGlnGlnGlnGly 107
QY      298  TTGACACACTGCCAGGTTAAAGATGCAGATGCGGACGAGTACAGTGCAGTCAGAT 357
      :      :      :      :      :      :      :      :      :      :
Db      108  GlnAspSerThrSerLysAlaAspLysProLysValAspLysGlnGlnGlnGlnGlu 127
QY      358  GAAGTTGATGAAGAT 372
      :      :      :      :      :      :      :      :      :      :
Db      128  Asp---AspLysAsp 131

RESULT 8
US-09-833-790-252
: Sequence 252, Application US/09833790
: Patent No. US20020068288A1
: GENERAL INFORMATION:
: APPLICANT: Lodes, Michael J.
: APPLICANT: Wang, Tonglong
: APPLICANT: Secrist, Heather
: APPLICANT: Mohamath, Radoch
: APPLICANT: Indrias, Carol Y.
: APPLICANT: Fan, Liqun
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
: TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
: FILE REFERENCE: 210121.512
: CURRENT APPLICATION NUMBER: US/09/833.790
: CURRENT FILING DATE: 2001-04-11
: NUMBER OF SEQ ID NOS: 440
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 252
: LENGTH: 156
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-833-790-252

Alignment Scores:
Pred. No.: 2.4      Length: 156
Score: 71.50      Matches: 26
Percent Similarity: 39.17%      Conservative: 21
Best Local Similarity: 21.67%      Mismatches: 50
Query Match: 10.61%      Indels: 23
      Gaps: 6
      DB: 10

US-09-936-737a-1 (1-375) x US-09-833-790-252 (1-156)
QY      61  TCAGAGAACGTGA-----GATGTTGACGCTTTTACGCCAACAAGAAA 105
      :      :      :      :      :      :      :      :      :      :
Db      29  AsnGluGlnGlnGlnLysLeuLysLysSerCysThrLeuTy-ValGlyAsnLeuSer 48
QY      106  TATACAGACTTCGATAAATCT-----TTTAAAGAGTCCTCGAT----- 144
      :      :      :      :      :      :      :      :      :      :
Db      49  PheTrpThrThrGlnGlnGlnIleTyrgLulePheSerLysSerGlyAspLysLys 68
QY      145  -----CTTGACGAATCCAAAAAACAATGTTTCAAGACGAGTACTGCTACATC 192
      :      :      :      :      :      :      :      :      :      :
Db      69  IleIleMetGlyLeuAspLysMetLysLysThrAla-----CysGlyPheCysPheVal 86
QY      193  GTTTTGAAGACACGCGTCAACAAGATGTTACTACATGTCGTTGATGTCGAAGATTA 252
      :      :      :      :      :      :      :      :      :      :
Db      87  GlnTyTrpSerArgAlaAspAlaGlnAsnAlaMetArgTrpLysGlnGlnGlnGlnGln 106
QY      253  GACCAAGAAAAATTTGTTGTCGACGAAAACTTCACGAAAAATTTATTTGACAGCTGCGAG 312
      :      :      :      :      :      :      :      :      :      :
Db      107  AspAspArg-----IleIleArgThrAspTrpAspAlaGlyPheLys-----Gln 121
QY      313  GGTAAAGATGCAGTAAATGCGGACGATACAGGTGACAGTCAATGATGATGAAGAT 372

```

```

Db      122  GlyArgGlnTrpCylArgGlyArgSerGlyGlnValArgAspGlnTrpArgGlnAsp 141
      :      :      :      :      :      :      :      :      :      :
RESULT 9
US-09-821-839-2
: Sequence 2, Application US/09821839
: Patent No. US20020129407A1
: GENERAL INFORMATION:
: APPLICANT: Ma, Hong
: APPLICANT: Ma, Hong
: TITLE OF INVENTION: Plant Gene Required for Male Meiosis
: FILE REFERENCE: Psu-0020
: CURRENT APPLICATION NUMBER: US/09/821.839
: CURRENT FILING DATE: 2001-03-29
: PRIOR APPLICATION NUMBER: 60/193,523
: PRIOR FILING DATE: 2000-03-31
: NUMBER OF SEQ ID NOS: 7
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 2
: LENGTH: 578
: TYPE: PRT
: ORGANISM: Arabidopsis thaliana
US-09-821-839-2

Alignment Scores:
Pred. No.: 4.51      Length: 578
Score: 70.50      Matches: 24
Percent Similarity: 37.61%      Conservative: 20
Best Local Similarity: 20.51%      Mismatches: 52
Query Match: 10.46%      Indels: 21
      Gaps: 4
      DB: 10

US-09-936-737a-1 (1-375) x US-09-821-839-2 (1-578)
QY      58  TCTTGAAGAAGACGCGAGATGTTGACGCTTTTACGCCAACAAGAAATATACAGCTTC 117
      :      :      :      :      :      :      :      :      :      :
Db      265  SerSerGlnGlnArgSerGlnIleTySerGlnTrpSerAspLysSerAspLys 284
QY      118  GATTAATCT-----TTTAAAGAGTCTCTGATCTTGACGAATCCAAAAACA----- 165
      :      :      :      :      :      :      :      :      :      :
Db      285  ThrProSerIlePhePheAspSerLysSerGlnPheSerGlnLysSerSerAspSer 304
QY      166  -----TGTTTCAAGACGAGTACTGCTACATCGTT 195
      :      :      :      :      :      :      :      :      :      :
Db      305  ProLysSerHisSerArgSerLeuTyrgLuleGlnPheLysGlnGlnPheCysArgSerThr 324
QY      196  TTTGAAGACACGCGTCAACAAGATGTTACTACAAATGTC-----GTT 237
      :      :      :      :      :      :      :      :      :      :
Db      325  IleProAsnAspPheGlySerSerCysGlnGlnGlnIleHisSerGlnLeuLeuArgPhe 344
QY      238  GATGTCGAAGAGTTAGACCAACAAGAAATTTGTTGTCGACGAAAACTTCACGAAAAAT--- 294
      :      :      :      :      :      :      :      :      :      :
Db      345  AspAspGlnGlnValGlnGlnSerTyrgLuleArgLeuArgGlnArgGlnArgSerHisAla 364
QY      295  TATTTGACAGACTCGAGGAGGTAAAGATGCAGGTTAATGCGGACGATACAGGT 345
      :      :      :      :      :      :      :      :      :      :
Db      365  TyrMetArgAspCysAlaLysAlaTyrgLysSerArgMetAspAsnThrGly 381

RESULT 10
US-09-815-242-11522
: Sequence 11522, Application US/09815242
: Patent No. US20020061569A1
: GENERAL INFORMATION:
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Karl L.
: APPLICANT: Zyskind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Essential Genes In
: TITLE OF INVENTION: Prokaryotes
: FILE REFERENCE: ELITRA.011A

```

```

? CLASSIFICATION : <Unknown>
? PRIOR APPLICATION DATA :
? APPLICATION NUMBER : 08/961,083
? FILING DATE : <Unknown>
? ATTORNEY/AGENT INFORMATION :
? NAME : Brookes, A. Anders
? REGISTRATION NUMBER : 36,373
? REFERENCE/DOCKET NUMBER : PB340P2
? TELECOMMUNICATION INFORMATION :
? TELEPHONE : (301) 309-8504
? TELEFAX : (301) 309-8512
? INFORMATION FOR SEQ ID NO: 200 :
? SEQUENCE CHARACTERISTICS :
? LENGTH: 674 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: protein
? SEQUENCE DESCRIPTION: SEQ ID NO: 200 :
US-09-765-272-200
Alignment Scores:

```

US-09-936-737A-1 (1-375) x US-09-765-272-200 (1-674)	
Freq. NO.:	6.16
Score:	69.50
Percent Similarity:	39.62%
Best Local Similarity:	23.58%
Query Match:	10.31%
DB:	10
Gaps:	5
Length:	67
Matches:	25
Conservative:	17
Mismatches:	35
Indels:	29

05 03-930-13/A-1 (1-3/5) x 05-09-765-212-200 (1-674,
07 82 TGGACGTTTACGCCGACAGAAAATATACAGACCTTCGATAA

US-09-765-272-200
Sequence 200, Application US/09765272
Patent No. US20020061545A1
GENERAL INFORMATION:
ADDENDUM: Chd 4

ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Kay West Avenue

CITY: Rockville
STATE: Maryland

STATE: Maryland
COUNTRY: USA

ZIP: 20850
COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.22
SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/765,272
FILING DATE: 22-Jan-2001

334 GCAGGTACGGGTGACGAG 351
QY
409 AlaTyrPheGlyGlnGlu 414
Db

RESULT 12
US-09-764-860-558
; Sequence 558, Application US/09764860
; Patent No. US20020094953A1
GENERAL INFORMATION:

APPLICANT: Rosen et al
TITLE OF INVENTION:

FILE REFERENCE: PC008

CURRENT APPLICATION NUMBER: US/09/764,860
CURRENT FILING DATE: 2001-01-17

Prior application data removed - consult PALM or file wrapper
NUMBER OF SEO ID NOS: 1198

SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 558

LENGTH: 72

ORGANISM: Homo sapiens

US-09-764-860-558

Alignment Scores:
Pred. No.: 4.36 Length: 72
Score: 68.50 Matches: 18
Percent Similarity: 51.67% Conservative: 13
Best Local Similarity: 30.00% Mismatches: 18
Query Match: 10.52% Indels: 11
DB: 10 Gaps: 2

US-09-936-737a-1 (1-375) x US-09-764-860-558 (1-72)

QY 324 TGCATCTTACCTCGAGTC-----TGTCAATTAATTTTC 289
Db 6 CysValpHeuGlnAlaLysThrThrThraSlnYsProHisCysArgIleGluAsn 25

QY 288 CGTAAATTTTC-----GTGACACAAATTTTCTTGCTTACCTCCACCATC 238
Db 26 ArgSerCysPheCysTyrGlyLeuArgValPheTyrIleLeuLeuMetIleLeu 45

QY 237 AACACATTTGATTAACATCTCTGACCGCTGTCTCAAAAACATGATGACATCTC 178
Db 46 AsnGlnMetLeuLeuThrPheLeuValAspCysArgTyrSerThrAsnIleAlaValLeu 65

RESULT 13
US-09-738-626-6537
; Sequence 6537, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738, 626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6537
; LENGTH: 206
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6537

Alignment Scores:
Pred. No.: 5.82 Length: 206
Score: 68.50 Matches: 25
Percent Similarity: 54.32% Conservative: 19
Best Local Similarity: 30.86% Mismatches: 26
Query Match: 10.16% Indels: 11
DB: 9 Gaps: 3

US-09-936-737a-1 (1-375) x US-09-738-626-6537 (1-206)

QY 134 AGTCCTGTGATCTTGAAGATGCAAAAAACATGTTTCAAGACGAGTACTGCTATACGC 193
Db 129 SerProGlnIleAlaThrAsnLeuArgLysPheLeuSerClnProProSerSerIleSer 148

QY 194 TTTTGAAGACAGCGTCAACAGATGTTACTACCAATGCTTGATGGGAAGATTAG 253
Db 149 AspleuProThrArgLys---ArgAspValLeuThrLeuLeuHisLeuGlyLysSer--- 166

QY 254 ACCAAGAAAAATTTGTTTCGACGAAACTTCACGGAATATTATTGACAGCTCGAG 313
Db 167 -----AsnAlaAspIleAlaIleGluLeuLysSerThrValSer 180

QY 314 GTAAGATGACGTA-----ATGCGCAGGTACAGTGCAGTCAGATGAAGTTG 364
Db 181 ValLysLysSerValSerArgLeuMetGlnArgPheGlnValSerSerArgLeuLeu 200

QY 365 ATG 367
Db 201 Val 201

RESULT 14
US-09-993-999-7
; Sequence 7, Application US/09993999
; Patent No. US20020110891A1
; GENERAL INFORMATION:
; APPLICANT: Ho, I-Cheng
; APPLICANT: Arm, Jonathan P.
; APPLICANT: Austen, K. Frank
; APPLICANT: Glincher, Laurie H.
; TITLE OF INVENTION: Phospholipase A2 Group Preferentially
; FILE REFERENCE: HU1-046
; CURRENT APPLICATION NUMBER: US/09/993, 999
; CURRENT FILING DATE: 2001-11-06
; PRIOR APPLICATION NUMBER: 60/246, 316
; PRIOR FILING DATE: 2000-11-06
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 146
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-993-999-7

Alignment Scores:
Pred. No.: 6.06 Length: 146
Score: 68.00 Matches: 15
Percent Similarity: 40.00% Conservative: 9
Best Local Similarity: 25.00% Mismatches: 18
Query Match: 10.09% Indels: 18
DB: 10 Gaps: 2

US-09-936-737a-1 (1-375) x US-09-993-999-7 (1-146)

QY 142 GATCTTGAAGATGCAAAAAACATGTTTCAAGACGAGTACTGCTATACGTTTGA 201
Db 62 AspleuAspArgCys-----CysGlnThrHisAspHisCysTyr----- 74

QY 202 GACACGGTCAACAGAGATGTTACTACAAATGCTTGATGGTGAAGATTAGACCAAGA 261
Db 75 -----SerGlnAlaLysLysLeuGluSerCys 83

QY 262 AAATTTGTTGCGACGAAACTTCACGGAATATTATTGACAGCTCGAGGTAAGAT 321
Db 84 LysPheLeuIleAspAsnProTyrThrAsnThrTyrSerTyrSerCysSerLysSerGlu 103

RESULT 15
US-09-801-368-426
; Sequence 426, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Call, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John

Run on: February 24, 2003, 20:43:39 ; search time 232 seconds (without alignments)

3640.083 Million cell updates/sec

Title: US-09-936-737A-1

Perfect score: 3/5
Sequence: 1 atgaagtatctctgattc.....atgaagtgatgaagattaa 375

Scoring table: IDENTITY_Nuc
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing:	Minimum Match	0%
	Maximum Match	100%

Listing first 45 summaries

Database :

- 1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT: *
- 2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT: *
- 3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT: *
- 4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT: *
- 5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT: *
- 6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT: *
- 7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT: *
- 8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT: *
- 9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT: *
- 10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT: *
- 11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT: *
- 12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT: *
- 13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT: *
- 14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT: *
- 15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT: *
- 16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT: *
- 17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT: *
- 18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT: *
- 19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT: *
- 20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT: *
- 21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT: *
- 22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT: *
- 23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT: *
- 24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT: *

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	375	100.0	375	21	AAA28180	Platelet binding
2	45.8	12.2	49999	20	AAZ23891	Murine LOBO genom
3	45.8	12.2	49999	20	AAZ23891	Murine LOBO genom
4	45.2	12.1	15935	23	ABJ06114	Drosophila melano
5	44.4	11.8	2438	24	ABO54916	Human ovarian ant
6	41	10.9	372	23	ABO54542	DNA encoding nove
7	40.2	10.7	3211	18	AAAT89346	Human p160 CDNA 1
8	40.2	10.7	3901	18	AAAT89345	Human p160 CDNA 1
9	39.6	10.6	766	22	ABA49279	Human breast cell

C	10	39.6	10.6	766	22	ABA67193	Human foetal liver
C	11	39.6	10.6	766	22	ABA54288	Probe #12754 for g
C	12	39.6	10.6	766	22	AA122114	Probe #12047 for g
C	13	39.6	10.6	766	22	AA147409	Probe #16095 used t
C	14	39.6	10.6	766	22	AA107812	Probe #7803 used t
C	15	39.6	10.6	1944	22	ABA44123	Human breast cell
C	16	39.6	10.6	1944	22	ABA54575	Human foetal liver
C	17	39.6	10.6	1944	22	ABA24559	Probe #2825 for ge
C	18	39.6	10.6	1944	22	AA134331	Probe #2807 for ge
C	19	39.6	10.6	1944	22	AA102791	Probe #2917 used t
C	20	39.6	10.6	3489	21	AAA30290	Probe #2782 used t
C	21	39.6	10.6	3489	22	AAAF82901	Kaposi's sarcoma-a
C	22	39.6	10.6	3489	22	ABA93487	Nucleotide sequenc
C	23	39.6	10.6	32207	20	AAV73805	Kaposi's sarcoma-a
C	24	39.6	10.6	137507	19	AAV19941	Kaposi's sarcoma-a
C	25	39.6	10.5	354	23	AAAS9067	KSHV LTR DNA (nucl
C	26	39.4	10.5	654	23	AAAS6953	KSHV long unique C
C	27	39.4	10.5	654	23	AAAS71153	DNA encoding novel
C	28	39.4	10.5	654	23	AAAS75457	DNA encoding novel
C	29	39.4	10.5	654	23	AAAS75455	DNA encoding novel
C	30	39.2	10.5	234	20	AAAS75460	Arabidopsis thalia
C	31	39	10.4	903	20	AAAS68550	DNA encoding novel
C	32	39	10.4	963	23	AAAC47655	DNA encoding novel
C	33	38.6	10.3	1448	21	AAAS90738	listeria monocytoc
C	34	38.4	10.2	1072	24	ABA03034	DNA encoding novel
C	35	37.8	10.1	2944528	23	AAAS64269	Plasmodium falcpa
C	36	37.6	10.0	1622	23	AAAS70213	Shrimp white spot
C	37	37.4	10.0	2658	21	AAAH62819	Shrimp white spot
C	38	37.4	10.0	3543	22	AAAB66889	Human breast cell
C	39	37.4	10.0	305107	22	ABA51452	Human foetal liver
C	40	37.2	9.9	315	22	ABA51452	Human breast cell
C	41	37.2	9.9	315	22	ABA65470	Human foetal liver
C	42	37.2	9.9	315	22	ABA35403	Probe #14869 for g
C	43	37.2	9.9	315	22	AAK17735	Human brain expres
C	44	37.2	9.9	315	22	AAK43555	Human bone marrow
C	45	37.2	9.9	315	22	AA124342	Probe #14275 for g

ALIGNMENTS

RESULT 1

ID	AAA28180	standard;	CDNA;	375	BP
Accession	AAA28180	standard;	CDNA;	375	BP

AC AAA28180;

DT 29-JAN-2001 (first entry)

aa Platelet binding inhibitor protein Saratin encoding cDNA sequence
DE

collagen-dependent platelet adhesion inhibition;

posterior capsule opacification; SS
KW

OS Hirudo medicinalis

Key	Location/Qualifiers
FH	

FT	CDS	64...
FT	CDS	64...

/product= "Saratin"

ET 227

PN WO2000056885-A1

PD 28-SEP-2000

PF 10-MAR-2000; 2000WO-EP02117

PR 18-MAR-1999; 99EP-0105530

XX

PA (MERE) MERCK PATENT GMBH

Same

Wed Feb 26 15:58:53 2003

us-09-936-737a-1.rng

AA23896 standard; DNA; 49999 BP.
 AA23896:
 25-JAN-2000 (first entry)
 Murine LOBO homologue genomic DNA fragment 2.
 LOBO: long bones; bone development; bone extension; skull; osteopathic;
 diagnostic; pharmaceutical; gene therapy; transgenic animal; disease;
 spondyloepiphyseal dysplasia; achondroplasia; murine; ds.
 Mus musculus.
 WO950284-A2.
 07-OCT-1999.
 26-MAR-1999: 99NO-EP02055.
 27-MAR-1998: 98DE-1013799.
 (ROSE/) ROSENTHAL A.
 Rosenthal A, Rump A, Hess J, Aigner T, Wirth T;
 WPI: 1999-601320/51.
 Nucleic acids encoding proteins which influence bone development,
 useful for treating and studying bone disorders -
 Example 3: Page 161-189; 391pp; German.
 This invention describes novel nucleic acids (I; designated LOBO (long
 bones)) encoding proteins influencing bone development in mammals. The
 proteins of the invention reduce and/or inactivate bone extension (i.e.
 development), with exception of the skull and have osteopathic activity.
 The nucleic acid molecules, proteins and antibodies can be used in
 diagnosis or pharmaceutical compounds e.g. for gene therapy. The methods
 and nucleic acid molecules, etc. are useful for production of transgenic
 animals, especially a transgenic mouse for the study of diseases
 associated with bone development, e.g. spondyloepiphyseal dysplasia and
 achondroplasia. This sequence encodes the murine LOBO protein described
 in the method of the invention.
 Sequence 49999 BP: 13135 A; 11787 C; 10868 G; 14209 T; 0 other;
 Query Match 12.2%; Score 45.8; DB 20; Length 49999;
 Best Local Similarity 49.8%; Pred. No. 0.011;
 Matches 116; Conservative 0; Mismatches 117; Indels 0; Gaps 0;
 Oy 143 ATCTTGACGAATGCAAAAACATGTTTCAAGCGAGTACTGCTACATGTTTGAAG 202
 Db 9318 ATTTTGAAGAAAAAATCTATATTACAGATTCACACTGTTATATAGAAAGAAAG 9377
 Oy 203 ACACGCTCAACAGAGATGTTACTACATGCTGTTGATGTCAGAGTTAGACCAAGAA 262
 Db 9378 AAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 9437
 Oy 263 AATTGTTGTCGACGAAAATCTTCAAGAAAATTTATTGACACACGCGGAGGTAAAGATG 322
 Db 9438 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 9497
 Oy 323 CAGGTAAATCGGCGAGTACAGTGCAGTGCAGTGAAGTTAA 375
 Db 9498 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 9550

RESULT 4
 ABL06114/C
 ID ABL06114 standard; cDNA: 15935 BP.
 XX
 AC ABL06114:

26-MAR-2002 (first entry)
 Drosophila melanogaster expressed polynucleotide SEQ ID NO 12824.
 DE Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.
 OS Drosophila melanogaster.
 WO200171042-A2.
 27-SEP-2001.
 23-MAR-2001; 2001WO-US09231.
 PF 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 PA (PEKE) PE CORP NY.
 PI Venter JC, Adams M, Li PWD, Myers EW.
 P-PSDB: ABB62011.
 WPI: 2001-656860/75.
 DR New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PI interactions -
 PS Claim 1: SEQ ID NO 12824; 21pp + Sequence listing; English.
 The invention relates to an isolated nucleic acid detection reagent
 capable of detecting 1000 or more genes from Drosophila. The invention is
 useful in developmental biology and in elucidating cell signalling and
 cell-cell interactions in higher eukaryotes for the development of
 insecticides, therapeutics and pharmaceutical drugs. The invention
 discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
 sequences (AB101840-AB116175) and the encoded proteins
 CC (AB57737-AB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.
 SO Sequence 15935 BP; 4219 A; 3351 C; 3394 G; 4971 T; 0 other;
 Query Match 12.1%; Score 45.2; DB 23; Length 15935;
 Best Local Similarity 49.6%; Pred. No. 0.011;
 Matches 116; Conservative 0; Mismatches 118; Indels 0; Gaps 0;
 Oy 141 TGATCTTGACGAATGCAAAAACATGTTTCAAGCGAGTACTGCTACATGTTTGA 200
 Db 10250 TAAATGATGGGACGACGATGACGAGATGACAAATGAGAGACAGACAGATTAACGA 260
 Oy 201 AGACAGGCTCAACAGAGATGTTACTACATGCTGTTGATGTCAGAGTTAGACCAAGAA 10131
 Db 10190 AGAAGATTAATTAATTAATGACGAGAGATGACGAGAGAGAGAGAGAGAGAGAGAG 320
 Oy 261 AAAATTTGTTGTCGACGAAAATCTTCAAGAAAATTTATTGACACACGCGGAGGTAAAGATG 10071
 Db 10130 CAATTAATGATTAATGACGAGAGATGACGAGAGAGAGAGAGAGAGAGAGAGAG 374
 Oy 321 TGCAGTAAATCGGCGAGTACAGTGCAGTGCAGTGAAGTTAA 10017
 Db 10070 AGAAGATTAATTAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 10017

RESULT 5
 AB054916
 ID AB054916 standard; cDNA: 2438 BP.
 XX
 AC AB054916:

CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at http://wipo.int/pub/publ/publseq_sequences.
 CC
 XX Sequence 372 BP: 194 A; 27 C; 134 G; 17 T; 0 other;
 SQ
 Query Match 10.9%; Score 41; DB 23; Length 372;
 Best Local Similarity 48.9%; Pred. No. 0.043; Indels 0; Gaps 0;
 Matches 110; Conservative 0; Mismatches 115;
 OY 151 GAATGCAAAAACATGTTTCAAGACGAGTACTGCTACATGCTTTTGAAGACACGGTC 210
 DB 31 GAATCCCTTAACGAGGATTCGAGACCTACCTAGCGCAACCTACCTGTA 90
 OY 211 AACAGGATGTTTACTACATGCTGATGCTGAGAGAGTACCAAGAAAATTTGTT 270
 DB 91 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 150
 OY 271 GTGACGAAACTTCACGGAATTTATTTTACAGACTGCGAGGTAAAGTGCAGTAAT 330
 DB 151 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 210
 OY 331 GCGGACGATGACGAGTACGAGTACGAGTACGAGTACGAGTACGAGTACGAG 375
 DB 211 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 255
 RESULT 7
 AAT89346
 ID AAT89346 standard; cDNA: 3211 BP.
 XX
 AC AAT89346;
 XX
 DT 11-MAR-1998 (first entry)
 XX
 DE Human p160 cDNA 160.2.
 XX
 KW p160; p62; cytoplasmic; T cell; B cell; development; activation;
 KW modulation; cellular response; cell proliferation; autoimmune disease;
 KW p56-lck; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT 439..3156
 FT /*tag= a
 FT /product= p160
 FT /note= "160.2"
 XX
 PN WO9722255-A1.
 XX
 PD 26-JUN-1997.
 XX
 PF 11-DEC-1996; 96WO-US19944.
 XX
 PR 19-DEC-1995; 95US-0574959.
 XX
 PA (DAND) DANA FARBER CANCER INST INC.
 XX
 PI Young I, Shin J, Strominger JL, Vadamudi RK;
 XX
 DR WPI: 1997-341351/31.
 XX
 PT cDNA encoding p62 and p160 and corresponding proteins - used in the
 PT treatment of autoimmune disease and for T and B cell proliferation,
 PI e.g. for treatment of tumors
 XX
 PS Claim 82; Fig 10; 175pp; English.
 XX

CC This cDNA sequence encodes a novel p160 (160.2) which is capable
 CC of activating transcription of a variety of genes upon activation of p62
 CC and is capable of binding to the p62/p56lck complex to modulate lck
 CC function in a manner similar to p62. The genes transcribed in response to
 CC p160 activation likely include those of which are involved in T or B cell
 CC development/differentiation, T or B cell activation or production of T or
 CC B cell specific factors e.g. lymphokines or antibodies. This p160
 CC polypeptide is also a substrate for serine/threonine kinase activity.
 CC p160 polypeptides can modulate degradation of cellular proteins e.g. cell
 CC cycle regulatory proteins stimulating expression of cell cycle dependent
 CC kinase inhibitors and arresting cell cycle progression at specific
 CC boundaries to thereby modulate cell proliferation. As p160 boosts B cell
 CC response it may be used to treat disorders where this is beneficial, e.g.
 CC infections by pathogenic microorganisms. p160 can be used to expand T
 CC cell populations for treating infectious diseases or cancer and p160
 CC inhibitors could reduce B or T cell responses and may be used to treat a
 CC variety of autoimmune diseases, e.g. diabetes mellitus, arthritis,
 CC multiple sclerosis allergic reactions, Crohn's diseases etc.
 CC
 XX Sequence 3211 BP: 649 A; 990 C; 948 G; 624 T; 0 other;
 SQ
 Query Match 10.7%; Score 40.2; DB 18; Length 3211;
 Best Local Similarity 52.0%; Pred. No. 0.17; Indels 0; Gaps 0;
 Matches 90; Conservative 0; Mismatches 83;
 OY 199 GAAGACAGCGTCAACAAGAGATGTTACTACATGCTGTTGATGTTGAGAGAGTACACCA 258
 DB 2458 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2517
 OY 259 GAAAAATTTGTTGTCGAGGAAACTTCACGAAATTTATTTGACAGACTGCGAGGTTAA 318
 DB 2518 GAATATTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2577
 OY 319 GATGAGGTTAATGCGGAGGTACGAGTACGAGTACGAGTACGAGTACGAGTACGAG 371
 DB 2578 GAGGTGAGTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2630
 RESULT 8
 AAT89345
 ID AAT89345 standard; cDNA: 3901 BP.
 XX
 AC AAT89345;
 XX
 DT 11-MAR-1998 (first entry)
 XX
 DE Human p160 cDNA 160.1.
 XX
 KW p160; p62; cytoplasmic; T cell; B cell; development; activation;
 KW modulation; cellular response; cell proliferation; autoimmune disease;
 KW p56-lck; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT 439..3846
 FT /*tag= a
 FT /product= p160
 FT /note= "160.1"
 XX
 PN WO9722255-A1.
 XX
 PD 26-JUN-1997.
 XX
 PF 11-DEC-1996; 96WO-US19944.
 XX
 PR 19-DEC-1995; 95US-0574959.
 XX
 PA (DAND) DANA FARBER CANCER INST INC.
 XX
 PI Young I, Shin J, Strominger JL, Vadamudi RK;
 XX
 DR WPI: 1997-341351/31.
 XX

DR P-PSDB; AAW31185.

XX CDNA encoding p62 and p160 and corresponding proteins - used in the
PT treatment of autoimmune disease and for T and B cell proliferation,
PT e.g. for treatment of tumours
XX

PS Claim 82; Fig 8; 175pp; English.

CC This cDNA sequence encodes a novel p160 (160.1) which is capable
CC of activating transcription of a variety of genes upon activation of p62
CC and is capable of binding to the p62/p56lck complex to modulate lck
CC function in a manner similar to p62. The genes transcribed in response to
CC p160 activation likely include those of which are involved in T or B cell
CC development/differentiation, T or B cell activation or production of T or
CC B cell specific factors e.g. lymphokines or antibodies. This p160
CC polypeptide is also a substrate for serine/threonine kinase activity.
CC p160 polypeptides can modulate degradation of cellular proteins e.g. cell
CC cycle regulatory proteins stimulating expression of cell cycle dependent
CC kinase inhibitors and arresting cell cycle progression at specific
CC boundaries to thereby modulate cell proliferation. As p160 boosts B cell
CC response it may be used to treat disorders where this is beneficial, e.g.
CC infections by pathogenic microorganisms. p160 can be used to expand T
CC cell populations for treating infectious diseases or cancer and p160
CC inhibitors could reduce B or T cell responses and may be used to treat a
CC variety of autoimmune diseases, e.g. diabetes mellitus, arthritis,
CC multiple sclerosis allergic reactions, Crohn's diseases etc.

SQ Sequence 3901 BP; 781 A; 1183 C; 1128 G; 809 T; 0 other;

Query Match Best Local Similarity 10.7%; Score 40.2; DB 18; Length 3901;

Matches 90; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 199 GAAGACAGGTCAACAGGANTGTACTACATGTCGTTGANTGTGAGAGTTAGACCA 258
DB 3148 GAAGAGAGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3207
QY 259 GAAATATTTGTCGACGAAACTTCACGAAATTTATTTGACGACCTCGAGGTAA 318
DB 3208 GAAATATTTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3267
QY 319 GATCAGGTAATGCGGAGGTACAGGTACAGGTACAGGTACAGGTACAGGTACAG 371
DB 3268 GAAGGTGAGTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3320

RESULT 9

ABAA9279/C
ID ABAA9279 standard; DNA; 766 BP.

XX ABAA9279;

AC 01-FEB-2002 (first entry)

DE Human breast cell single exon nucleic acid probe #7974.

KW Human; microarray; single exon probe; gene expression; breast;

XX disease; cancer; ss.

OS Homo sapiens.

XX WO200157271-A2.

PD 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00662.

PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-496933/54.

PT New spatially-addressable set of single exon nucleic acid probes,
PT useful for measuring gene expression in sample derived from human
PT breast, comprises number of single exon nucleic acid probes

PS Claim 4; SEQ ID NO 7974; 327pp + sequence listing; English.

CC The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and Bt 474 cells. The method involves contacting
CC the probes with a collection of detectably labelled nucleic acids
CC derived from mRNA of human breast, and then measuring the label
CC bound to each probe of the microarray. The probes are useful for
CC verifying the expression of regions of genomic DNA predicted to
CC encode proteins. They are useful for gene discovery, and for
CC determining predisposition and/or assessing the toxicity of chemical
CC agents on cells. The microarray of this invention presents a far greater
CC diversity of probes for measuring gene expression, with far less bias
CC than expressed sequence tag microarrays. The method is suitable for
CC rapid production of functional information from genomic sequence. The
CC present sequence is a single exon nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pcl_sequences.

SQ Sequence 766 BP; 190 A; 299 C; 36 G; 241 T; 0 other;

Query Match Best Local Similarity 10.6%; Score 39.6; DB 22; Length 766;

Matches 93; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 187 TACATCGTTTGAAGACAGCGTCAACAGGATGTTACTACATGTCGTTGANTGGA 246
DB 687 TCCACACTTGAAGACAGCGTCAACAGGATGTTACTACATGTCGTTGANTGGA 246
QY 247 GAGTTAGACCAAGAAATTTGTTGACGAAACTTCACGAAATTTATTTGACAGC 306
DB 627 GATGAGAGGATGATTAATGATGACATGATGATGATGATGATGATGATGATGAT 628
QY 307 TCGAGGTTAAAGATGCGAGTATGCGGACAGGTACAGGTACAGGTACAGGTAT 366
DB 567 GACGAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 568
QY 367 GA 368
DB 507 GA 506

RESULT 10

ABA67193/C
ID ABA67193 standard; DNA; 766 BP.

XX ABA67193;

AC 01-FEB-2002 (first entry)

DE Human foetal liver single exon nucleic acid probe #15498.

KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.

XX Homo sapiens.

OS WO200157271-A2.
XX PD 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00669.
 PF
 XX 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-483447/52.
 DR
 PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human fetal liver -
 XX
 PS Claim 4; SEQ ID NO 15498; 639pp + sequence listing; English.
 XX
 CC The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,
 CC measuring and displaying gene expression in samples derived from human
 CC fetal liver. The present sequence is a single exon nucleic acid
 CC probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX
 SQ Sequence 766 BP; 190 A; 299 C; 36 G; 241 T; 0 other;
 Query Match 10.6%; Score 39.6; DB 22; Length 766;
 Best Local Similarity 51.1%; Pred. No. 0.14; Indels 0; Gaps 0;
 Matches 93; Conservative 0; Mismatches 89; Indels 0; Gaps 0;
 QY 187 TACATCGTTTTGAAGACGCGTCAACAGGAATGTTACTACATGCTGTGATGTA 246
 Db 687 TCACACTTGAAGACCAACCATGAGATGATGATGATGAAGATGATGATGAT 628
 QY 247 GAGTTAGACCAAGAAAATTTGTTGACGAAAACCTTCACGAAAATTTTGCACAG 306
 Db 627 GATGGAGAGGATGATTAATATGATGACGATGATGATGATGATGATGATGATG 568
 QY 307 TGGGAGGTTAAAGATGAGGTATGGGCGAGTACAGGTACAGATGATGATGAT 366
 Db 567 GACGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 508
 QY 367 GA 368
 Db 507 GA 506
 RESULT 11
 ABA34288/C
 ID ABA34288 standard; DNA; 766 BP.
 AC ABA34288;
 XX
 DE 23-JAN-2002 (first entry)
 XX
 DE Probe #12754 for gene expression analysis in human heart cell sample.
 XX
 KW Human; gene expression; heart; microarray; vascular system; probe;
 KW cardiovascular disease; hypertension; cardiac arrhythmia;
 KW congenital heart disease; ss.
 XX
 OS Homo sapiens.
 XX
 XX MO200157274-A2.
 XX

PD 09-AUG-2001.
 PF
 XX 30-JAN-2001; 2001WO-US00666.
 PF
 XX 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-488899/53.
 DR
 PT Single exon nucleic acid probes for analyzing gene expression in human
 PT hearts -
 XX
 PS Claim 4; SEQ ID NO 12754; 530pp; English.
 XX
 CC The present invention relates to single exon nucleic acid probes for
 CC measuring human gene expression in a sample derived from human heart. The
 CC present sequence is one such probe. The probes may be used for
 CC predicting, measuring and displaying gene expression in samples derived
 CC from the human heart via microarrays. By measuring gene expression, the
 CC probes are useful for predicting, diagnosing, grading, staging,
 CC monitoring and prognosing diseases of the human heart and vascular system
 CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
 CC congenital heart disease.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX
 SQ Sequence 766 BP; 190 A; 299 C; 36 G; 241 T; 0 other;
 Query Match 10.6%; Score 39.6; DB 22; Length 766;
 Best Local Similarity 51.1%; Pred. No. 0.14; Indels 89; Gaps 0;
 Matches 93; Conservative 0; Mismatches 89; Indels 0; Gaps 0;
 QY 187 TACATCGTTTTGAAGACGCGTCAACAGGAATGTTACTACATGCTGTGATGTA 246
 Db 687 TCACACTTGAAGACCAACCATGAGATGATGATGATGATGAAGATGATGATGAT 628
 QY 247 GAGTTAGACCAAGAAAATTTGTTGACGAAAACCTTCACGAAAATTTTGCACAG 306
 Db 627 GATGGAGAGGATGATTAATATGATGACGATGATGATGATGATGATGATGATGAT 568
 QY 307 TGGGAGGTTAAAGATGAGGTATGGGCGAGTACAGGTACAGATGATGATGAT 366
 Db 567 GACGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 508
 QY 367 GA 368
 Db 507 GA 506
 RESULT 12
 AA122114/C
 ID AA122114 standard; DNA; 766 BP.
 AC AA122114;
 XX
 DE 12-OCT-2001 (first entry)
 XX
 DE Probe #12047 for gene expression analysis in human cervical cell sample.
 XX
 KW Probe; human; microarray; gene expression; cervical epithelial cell;
 KW cervical cancer; ss.
 XX
 OS Homo sapiens.
 XX

[illegible]

•
•
•
•

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2003, 00:54:10 ; Search time 46 Seconds
(without alignments)
2500.082 Million cell updates/sec

Title: US-09-936-737A-1

Perfect score: 375
Sequence: 1 atgaagattcttctgatttc.....atgaagttgatgaagattaa 375

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*

1: /cgn2_6/pdata/1/lna/5A.COMB.seq:*
2: /cgn2_6/pdata/1/lna/5B.COMB.seq:*
3: /cgn2_6/pdata/1/lna/6A.COMB.seq:*
4: /cgn2_6/pdata/1/lna/6B.COMB.seq:*
5: /cgn2_6/pdata/1/lna/PCITUS.COMB.seq:*
6: /cgn2_6/pdata/1/lna/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	42.2	11.3	7218	1	US-08-232-463-14
2	40.2	10.7	3211	2	US-08-574-959A-8
3	40.2	10.7	3211	4	US-09-357-014-8
4	40.2	10.7	3901	2	US-08-574-959A-6
5	40.2	10.7	3901	4	US-09-357-014-6
6	39.6	10.6	3489	2	US-08-728-323A-1
7	39.6	10.6	3489	4	US-09-298-568-1
8	39.6	10.6	32207	2	US-08-770-379-20
C 9	39.6	10.6	32207	4	US-08-757-669A-20
C 10	39.6	10.6	32207	4	US-09-230-371A-20
11	36	9.6	966	2	US-08-766-738-2
12	36	9.6	966	4	US-09-262-610-2
13	35	9.3	2694	2	US-08-867-941-2
14	35	9.3	2694	4	US-09-074-658-2
15	35	9.3	7650	2	US-08-867-941-1
16	35	9.3	7650	4	US-09-074-658-1
17	35	9.3	8920	2	US-08-446-855A-1
18	35	9.3	8920	4	US-09-150-741-1
19	34.8	9.3	5361	4	US-08-973-462-2
20	34.8	9.3	6152	4	US-08-973-462-1
21	34.6	9.2	2518	4	US-09-433-699-3
22	34.2	9.1	5852	1	US-07-867-106-2
23	34	9.1	759	1	US-08-466-603-4
24	34	9.1	759	1	US-08-314-503A-4
25	34	9.1	759	1	US-08-468-066-4
26	34	9.1	759	1	US-08-466-717-4
27	34	9.1	759	3	US-08-466-743-4

28	34	9.1	759	5	PCT-US95-12414-4	Sequence 4, Appl
29	34	9.1	980	1	US-08-466-603-3	Sequence 3, Appl
30	34	9.1	980	1	US-08-314-503A-3	Sequence 3, Appl
31	34	9.1	980	1	US-08-468-066-3	Sequence 3, Appl
32	34	9.1	980	2	US-08-466-717-3	Sequence 3, Appl
33	34	9.1	980	2	US-08-466-743-3	Sequence 3, Appl
34	34	9.1	980	5	PCT-US95-12414-3	Sequence 3, Appl
35	33.8	9.0	4518	4	US-08-961-527-121	Sequence 121, App
36	33.6	9.0	2295	1	US-08-375-300-3	Sequence 3, Appl
37	33.6	9.0	2295	1	US-09-177-431-3	Sequence 3, Appl
38	33.6	9.0	2295	5	PCT-US95-16930-3	Sequence 3, Appl
39	33.6	9.0	4080	1	US-08-375-300-1	Sequence 1, Appl
40	33.6	9.0	4080	3	US-09-177-431-1	Sequence 1, Appl
41	33.6	9.0	4080	5	PCT-US95-16930-1	Sequence 1, Appl
42	33.4	8.9	1137	4	US-09-134-001C-657	Sequence 657, App
43	33.4	8.9	1727	4	US-09-071-035-295	Sequence 295, App
44	33.4	8.9	1839	4	US-09-071-035-293	Sequence 293, App
45	33.2	8.9	4599	1	US-08-431-080-27	Sequence 27, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEFFELINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESS: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
; US-08-232-463-14
Query Match 11.3%; Score 42.2; DB 1; Length 7218;

Best Local Similarity	52.0%;	Pred. No.	0.0066;
Matches	90;	Conservative	0;
		Mismatches	83;
		Indels	0;
		Gaps	0;

199 GAAGACACGGTCAACAAGGATGTTACTACAATGTCGTTGATGGTGAAGAGTTAGACCAA 258

b 2458 GAAGGAAGGAGAAGAGAGAGGAACTTTGAGGAAGAGAGGATGAAGAG 2517

259 GAAAAATTTGTGTGTCGACGAAACTTCACGGAAATTAATTTGACAGACTTCGAGGTTAA 318

25 / 230

RESULT 3
S-09-357-014-8

Sequence 8, Application US/09357014
Patent No. 6291645

GENERAL INFORMATION:
APPLICANT: Jaekyoon Shin, Insil Young, Ratna K. Vadlamudi

TITLE OF INVENTION: p62 POLYPEPTIDES, RELATED POLYPEPTIDES

NUMBER OF SEQUENCES: 22

ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street Suite 510

CITY: Boston
STATE: Massachusetts

COONIKI: USA
ZIP: 02109-1875

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US 409/3557 014

PRIOR APPLICATION DATA:
FILING DATE: 19-JUL-1999

APPLICATION NUMBER: 08/3/4,939
FILING DATE: <Unknown>

NAME: Mandragouras, Amy E.

REFERENCE/DOCKET NUMBER: DFN-008

TELEPHONE: (617)227-7400
FAX: (617)227-5041

INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:

LENGTH: 3211 base pairs
TYPE: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear

FEATURE: 11.51 CDR:

LOCATION: 439.3157

S-09-357-014-8

Query Match	10.78;	Score 40.2;	DB 4;	Length 3211;
Best Local Similarity	52.08;	Pred NO	0	0066;

Matches	90;	Conservative	0;	Mismatches	83;	Indels	0;	Gaps	
---------	-----	--------------	----	------------	-----	--------	----	------	--

179 GAGGACACGGICAAACCAAGGAAIGIATCAGCAAGICGIIAGAGGAGGAAAGGIIAGAACCA 230

2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526

```

      : GENERAL INFORMATION:
      : APPLICANT: Jaekyoon Shin, Insil Jung, Ratna K. Vadlamudi
      :                and Jack L. Strominger
      : TITLE OF INVENTION: p62 POLYPEPTIDES, RELATED POLYPEPTIDES
      :                AND USES THEREFOR
      : NUMBER OF SEQUENCES: 22
      : CORRESPONDENCE ADDRESS:
      :   ADDRESSEE: LAHIVE & COCKFIELD
      :   STREET: 60 State Street, Suite 510
      :   CITY: Boston
      :   STATE: Massachusetts
      :   COUNTRY: USA
      :   ZIP: 02109-1875
      : COMPUTER READABLE FORM:
      :   MEDIUM TYPE: Floppy disk
      :   OPERATING SYSTEM: PC-DOS/M5-DOS
      :   SOFTWARE: PatentIn Release #1.0, Version #1.25
      : CURRENT APPLICATION DATA:
      :   APPLICATION NUMBER: US/09/357,014
      :   FILING DATE: 19-Jul-1999
      : PRIOR APPLICATION DATA:
      :   APPLICATION NUMBER: 08/574,959
      :   FILING DATE: <Unknown>
      : ATTORNEY/AGENT INFORMATION:
      :   NAME: Mandragouras, Amy E.
      :   REGISTRATION NUMBER: 36,207
      :   REFERENCE/DOCKET NUMBER: DFN-008
      : TELECOMMUNICATION INFORMATION:
      :   TELEPHONE: (617)227-7400
      :   TELEFAX: (617)227-5941
      : INFORMATION FOR SEQ ID NO: 6:
      :     SEQUENCE CHARACTERISTICS:
      :       LENGTH: 3901 base pairs
      :       TYPE: nucleic acid
      :       STRANDEDNESS: single
      :       TOPOLOGY: linear
      : MOLECULE TYPE: cDNA
      : FEATURE:
      :   NAME/KEY: CDS
      :   LOCATION: 439..3847
      : ? ? ? SEQUENCE DESCRIPTION: SEQ ID NO: 6:
      : US-09-357-014-6
    
```

Query Match 10.7% ; Score 40.2 ; DB 4 ; Length 3901;
Best Local Similarity 52.0%; Pred.No. 0.0073;
Matches 90; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

```

QY      199 GAAGACACGCGTCAACAAGAATGGTTACTACCATGTGCCTTGATTGGTGAAGATAAGCCAA 258
        |||||  |  |  |||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      3148 GAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCCTTTTAGAAGAGAGAGAGAGATGAAGAG 3207
QY      259 GAAAAATTTCGTGTGCAGCAAACCTTCACGSAANAATTATTTTGACAGACTGCCAGGGGTAAA 318
        |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      3208 GAATTATTTTGAAGAGAGAGAGAGAGAGAGAGAMAAGATTGTAGAGAACATTTTGAGGAAGA 3267
QY      319 GATGCAGCGTAATGCGGACGGGTACAGCGTGACAGCGCATGTAAGTTGATGAAGA 371
        ||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      3268 GAAGGTGCACTTAGAGAGAGAGAGAGAGAGAGAGATGAGAGAGAGAGAGA 3320
    
```

RESULTS
US-08-728-323A-1
Sequence 1, Application US/08728323A
Patent No. 5948676

GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: Immediate Early Protein From Kaposi's
Sarcoma-Associated Herpesvirus, DNA

```

; TITLE OF INVENTION: Encoding Same And Uses Thereof
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,323A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 0575/52268/JPW/NSC/SKS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-278-0400
; TELEFAX: 212-391-0525
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3489 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3489
; US-08-728-323A-1

Query Match          10.6%; Score 39.6; DB 2; Length 3489;
Best Local Similarity 47.2%; Pred. No. 0.011;
Matches 120; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

; 118 GATAAATCTTTAAGAACTCTGATCTTGACGAATGCAAAAAACATGTTCAAGAG 177
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 871 GAAGAAGCATCCAAAGATGAAAAAGATGTCGGAATAATAGCTGGCCGAGTAAAT 930
;
; QY 178 GAGTACTGCTACATCGTTTTTGAAGACACGCGTCAACAGAAATGTACTCAATGTCGT 237
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 931 GGGGACACAGAGATTAGCAAGAAATGTCAGGTTGACAAAGATGCAATGCAATTAAGAT 990
;
; QY 238 GATGCTGAAGATTAGACCAAGAAATTTGTTGTCGACGAAACTTCACGAAATTAAT 297
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 991 GATGAGAGAGACAGAGACAGATGAGAGAGACGAGAGATACCGAGAGGATGACAG 1050
;
; QY 298 TTGACAGACTCGAGGCTAAAGATGCGATTAATGCGCAGAGTACAGGTGACAGAT 357
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 1051 GAGGATGACGAGGATGACGAGAGATGACGAGAGATGACGAGAGATGACGAG 1110
;
; QY 358 GAAGTTGATGAAGA 371
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 1111 GAGGATGACGAGGA 1124

RESULT 7
US-09-298-568-1
; Sequence 1, Application US/09298568
; Patent No. 6322792
; GENERAL INFORMATION:
; APPLICANT: Kieff, Elliott D.
; APPLICANT: Ballestras, Mary E.
; APPLICANT: Kaye, Kenneth M.
; TITLE OF INVENTION: RHADINO VIRUS LANA ACTS IN TRANS ON A UNIT OF RHADINO
; TITLE OF INVENTION: VIRUS DNA TO MEDIATE EFFICIENT EPISOME PERSISTENCE
; FILE REFERENCE: 16412-10001R

```

```

; CURRENT APPLICATION NUMBER: US/09/298,568
; CURRENT FILING DATE: 1999-04-21
; EARLIER APPLICATION NUMBER: US 60/109,422
; EARLIER FILING DATE: 1998-11-19
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3489
; TYPE: DNA
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
; US-09-298-568-1

Query Match          10.6%; Score 39.6; DB 4; Length 3489;
Best Local Similarity 47.2%; Pred. No. 0.011;
Matches 120; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

; 118 GATAAATCTTTAAGAACTCTGATCTTGACGAATGCAAAAAACATGTTCAAGAG 177
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 871 GAAGAAGCATCCAAAGATGAAAAAGATGTCGGAATAATAGCTGGCCGAGTAAAT 930
;
; QY 178 GAGTACTGCTACATCGTTTTTGAAGACACGCGTCAACAGAAATGTACTCAATGTCGT 237
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 931 GGGGACACAGAGATTAGCAAGAAATGTCAGGTTGACAAAGATGCAATGCAATTAAGAT 990
;
; QY 238 GATGCTGAAGATTAGACCAAGAAATTTGTTGTCGACGAAACTTCACGAAATTAAT 297
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 991 GATGAGAGAGACAGAGACAGATGAGAGAGACGAGAGATGACGAGAGATGACGAG 1050
;
; QY 298 TTGACAGACTCGAGGCTAAAGATGCGATTAATGCGCAGAGTACAGGTGACAGAT 357
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 1051 GAGGATGACGAGGATGACGAGAGATGACGAGAGATGACGAGAGATGACGAGAGATGACGAG 1110
;
; QY 358 GAAGTTGATGAAGA 371
;   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 1111 GAGGATGACGAGGA 1124

RESULT 8
US-08-770-379-20/c
; Sequence 20, Application US/08770379
; Patent No. 5849564
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
; TITLE OF INVENTION: HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/770,379
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 52342
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525

```

; INFORMATION FOR SEQ ID NO: 20:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 32207 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; US-08-770-379-20

Query Match 10.6%; Score 39.6; DB 2; Length 32207;
 Best Local Similarity 47.2%; Pred. No. 0.03;
 Matches 120; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

QY 118 GATAAATCTTTAAGAAAGCTCTGATCTTGACGATGCAAAAAACATGTTTCAAGACG 177
 DB 21126 GAAGAAGCATCCAAAGATGAAAAAGATGTCGAAAAATATCAAGCTGGGAGAGATTAAT 21067
 QY 178 GAGTACTGCTACATCGTTTTTGAAGACACGGTCAACAGAAATGTTACTACATGTCGTT 237
 DB 21066 GGGGACACGACGATTTAGCAAGAAAGTCAGGTTGACAAAGATGACAAATTAAGGAT 21007
 QY 238 GATGTTGAAGAGTTAGACCAAGAAAAATTTGTTCCGACGAAAACTTCACGAAAAATTAAT 297
 DB 21006 GATGAGAGAGACGACGAGACAGATGAGAGAGAGAGATGACGAGAGATGACGAG 20947
 QY 298 TTGACAGACTGCGAGGTTAAAGATGACGATGATGCGCAGGTACAGGTGACGAGATCAGAT 357
 DB 20946 GAGGATGACGAGAGAGATGACGAGAGAGATGACGAGAGATGACGAGAGATGACGAG 20887
 QY 358 GAAGTTGATGAAGA 371
 DB 20886 GAGGATGACGAGGA 20873

RESULT 9
 US-08-757-669A-20/c
 ; Sequence 20, Application US/08757669A
 ; Patent No. 6183751

; GENERAL INFORMATION:
 ; APPLICANT: Chang, Yuan
 ; APPLICANT: Bohenzky, Roy A.
 ; APPLICANT: Russo, James J.
 ; APPLICANT: Edelman, Isidore S.
 ; APPLICANT: Moore, Patrick S.
 ; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
 ; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
 ; NUMBER OF SEQUENCES: 20
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Cooper & Dunham LLP
 ; STREET: 1185 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: U.S.A.
 ; ZIP: 10036
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/757,669A
 ; FILING DATE:
 ; CLASSIFICATION: 424
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: White, John P.
 ; REGISTRATION NUMBER: 28,678
 ; REFERENCE/DOCKET NUMBER: 45185-F
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 278-0400
 ; TELEFAX: (212) 391-0525
 ; INFORMATION FOR SEQ ID NO: 20:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 32207 base pairs

; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; US-08-757-669A-20

Query Match 10.6%; Score 39.6; DB 4; Length 32207;
 Best Local Similarity 47.2%; Pred. No. 0.03;
 Matches 120; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

QY 118 GATAAATCTTTAAGAAAGCTCTGATCTTGACGATGCAAAAAACATGTTTCAAGACG 177
 DB 21126 GAAGAAGCATCCAAAGATGAAAAAGATGTCGAAAAATATCAAGCTGGGAGAGATTAAT 21067
 QY 178 GAGTACTGCTACATCGTTTTTGAAGACACGGTCAACAGAAATGTTACTACATGTCGTT 237
 DB 21066 GGGGACACGACGATTTAGCAAGAAAGTCAGGTTGACAAAGATGACAAATTAAGGAT 21007
 QY 238 GATGTTGAAGAGTTAGACCAAGAAAAATTTGTTCCGACGAAAACTTCACGAAAAATTAAT 297
 DB 21006 GATGAGAGAGACGACGAGACAGATGAGAGAGAGATGACGAGAGATGACGAG 20947
 QY 298 TTGACAGACTGCGAGGTTAAAGATGACGATGATGCGCAGGTACAGGTGACGAGATCAGAT 357
 DB 20946 GAGGATGACGAGAGAGATGACGAGAGAGATGACGAGAGATGACGAGAGATGACGAG 20887
 QY 358 GAAGTTGATGAAGA 371
 DB 20886 GAGGATGACGAGGA 20873

RESULT 10
 US-09-230-371A-20/c
 ; Sequence 20, Application US/09230371A
 ; Patent No. 6348586

; GENERAL INFORMATION:
 ; APPLICANT: Chang, Yuan
 ; APPLICANT: Bohenzky, Roy A.
 ; APPLICANT: Russo, James J.
 ; APPLICANT: Edelman, Isidore S.
 ; APPLICANT: Moore, Patrick S.
 ; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS SEQUENCES AND
 ; TITLE OF INVENTION: USES THEREOF
 ; FILE REFERENCE: 45185-G-PCT-US
 ; CURRENT APPLICATION NUMBER: US/09/230,371A
 ; PRIOR FILING DATE: 1999-11-17
 ; PRIOR APPLICATION NUMBER: PCT/US97/13346
 ; NUMBER OF SEQ ID NOS: 30
 ; SOFTWARE: Patentin Ver. 2.0
 ; SEQ ID NO 20
 ; LENGTH: 32207
 ; TYPE: DNA
 ; ORGANISM: Kaposi's sarcoma-associated herpesvirus
 ; US-09-230-371A-20

Query Match 10.6%; Score 39.6; DB 4; Length 32207;
 Best Local Similarity 47.2%; Pred. No. 0.03;
 Matches 120; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

QY 118 GATAAATCTTTAAGAAAGCTCTGATCTTGACGATGCAAAAAACATGTTTCAAGACG 177
 DB 21126 GAAGAAGCATCCAAAGATGAAAAAGATGTCGAAAAATATCAAGCTGGGAGAGATTAAT 21067
 QY 178 GAGTACTGCTACATCGTTTTTGAAGACACGGTCAACAGAAATGTTACTACATGTCGTT 237
 DB 21066 GGGGACACGACGATTTAGCAAGAAAGTCAGGTTGACAAAGATGACAAATTAAGGAT 21007
 QY 238 GATGTTGAAGAGTTAGACCAAGAAAAATTTGTTCCGACGAAAACTTCACGAAAAATTAAT 297
 DB 21006 GATGAGAGAGACGACGAGACAGATGAGAGAGAGATGACGAGAGATGACGAGAGATGACGAG 20947
 QY 298 TTGACAGACTGCGAGGTTAAAGATGACGATGATGCGCAGGTACAGGTGACGAGATCAGAT 357

RESULT 13
US-08-867-941-2

RESULT 13

Sequence 2, Application US/08867941
Patent No. 5977337
GENERAL INFORMATION:
APPLICANT: Loosmore, Sheena M
APPLICANT: Du, Run-Pan
APPLICANT: Wang, Qiu-Jun
APPLICANT: Yang, Yan-Ping
APPLICANT: Klein, Michel H
TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA
NUMBER OF SEQUENCES: 67
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6th Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/867,941
FILING DATE: 03-JUN-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-681 MIS:jb
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1153
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2694 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-867-941-2

Query Match
Best Local Similarity 9.38; Score 35; DB 2; Length 2694;
Matches 105; Conservative 0; Mismatches 90; Indels 6; Gaps 1;

Db 172 AAGACGAGTACTGCTACATCGTTTGAAGACAGGTCACCAAGAATGTTACTACAT 231
Db 2017 AAAAAGGACAAAGTTATAGCAATATGAGAAACCATCAAGAAAAAGCCATCAAGAT 2076
QY 232 GTCGTTGATGTAAGAGTTAGACCAAGAAAAATTTGTCGACGCAAACTTCACGAA 291
Db 2077 TATCTGTAAACCGAAGACTTCAACCCACGAA-----GATGATGACGATGATTGACCGCA 2130
QY 292 AATTATTGACAGACTGCGAGGTAAGATGCGAGTATGCGGACAGTACAGGTGACGAG 351
Db 2131 TCTGATGATTGACCAAGATGATGATGACATGCGCATGATGATTGATGATCATCATAT 2190
QY 352 TCAGATGAAGTTGATGAAGAT 372
Db 2191 TCACAAGATGATGACGAGAT 2211

RESULT 14
US-09-074-658-2
Sequence 2, Application US/09074658
Patent No. 6184371
GENERAL INFORMATION:
APPLICANT: Loosmore, Sheena M
APPLICANT: Du, Run-Pan
APPLICANT: Wang, Qiu-Jun
APPLICANT: Yang, Yan-Ping
APPLICANT: Klein, Michel H
TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA

NUMBER OF SEQUENCES: 78
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6th Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/074,658
FILING DATE: 08-MAY-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-795
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1153
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2694 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-074-658-2

Query Match
Best Local Similarity 9.38; Score 35; DB 4; Length 2694;
Matches 105; Conservative 0; Mismatches 90; Indels 6; Gaps 1;

QY 172 AAGACGAGTACTGCTACATCGTTTGAAGACAGGTCACCAAGAATGTTACTACAT 231
Db 2017 AAAAAGGACAAAGTTATAGCAATATGAGAAACCATCAAGAAAAAGCCATCAAGAT 2076
QY 232 GTCGTTGATGTAAGAGTTAGACCAAGAAAAATTTGTCGACGCAAACTTCACGAA 291
Db 2077 TATCTGTAAACCGAAGACTTCAACCCACGAA-----GATGATGACGATGATTGACCGCA 2130
QY 292 AATTATTGACAGACTGCGAGGTAAGATGCGAGTATGCGGACAGTACAGGTGACGAG 351
Db 2131 TCTGATGATTGACCAAGATGATGATGACATGCGCATGATGATTGATGATCATCATAT 2190
QY 352 TCAGATGAAGTTGATGAAGAT 372
Db 2191 TCACAAGATGATGACGAGAT 2211

RESULT 15
US-08-867-941-1
Sequence 1, Application US/08867941
Patent No. 5977337
GENERAL INFORMATION:
APPLICANT: Loosmore, Sheena M
APPLICANT: Du, Run-Pan
APPLICANT: Wang, Qiu-Jun
APPLICANT: Yang, Yan-Ping
APPLICANT: Klein, Michel H
TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6th Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:

Wed Feb 26 15:58:54 2003

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/867,941
FILING DATE: 03-JUN-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-681 MIS:jb
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 7650 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-867-941-1

Query Match          9.3%; Score 35; DB 2; Length 7650;
Best Local Similarity 52.2%; Pred. No. 0.42;
Matches 105; Conservative 0; Mismatches 90; Indels 6; Gaps 1;

QY 172 AAGACGAGTACTGCTACATCGTTTGAAGACACGCGTCACACAGAAATGTTACTACAT 231
    ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2157 AAAAAGACAAAGTTATAGCAATATGAGAAACCATCAGAAAAAAGGCCATCAGAT 2216
    ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 232 GTGCTGATGTGAAGAGTTAGACCAAGAAAAATTTGTCGACGAAATCTTCACGGAA 291
    | ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2217 TATCTGTTAACCGAGACTTCACCCAGAA-----GATGATGACGATGTTGACCCGA 2270
    | ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 292 AATTATTGACAGACTCGAGGGTAAAGATCAGAGTAATCGGACAGTACAGGTGACGAG 351
    ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2271 TCTGATGATTCACAGATATGATGCATGCGATGATGATTTGATTGATCTGATGAT 2330
    ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 352 TCAGATGAGTTGATGATGAT 372
    ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2331 TCACAGATGATGACGAGAT 2351
    ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Search completed: February 25, 2003, 01:50:29
Job time : 71 secs

Gencore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2003, 01:05:25 ; Search time 61 seconds

(without alignments)
3452.701 Million cell updates/sec

Title: US-09-936-737a-1

Perfect score: 375
Sequence: 1 atgaagtattcttgatttc.....atgaagtgtatgaatttaa 375

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 442118 seqs, 280819700 residues

Total number of hits satisfying chosen parameters: 884236

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published_Applications_NA:*
1: /cgn2_6/prodata/1/pubpna/us07_PUBCOMB.seq:*
2: /cgn2_6/prodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/prodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/prodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/prodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/prodata/1/pubpna/PCITUS_PUBCOMB.seq:*
7: /cgn2_6/prodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/prodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/prodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/prodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/prodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/prodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/prodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	39.6	10.6	766	10 US-09-864-761-19608	Sequence 19608, A
2	39.6	10.6	1944	10 US-09-864-761-2825	Sequence 2825, Ap
3	37.2	9.9	315	10 US-09-864-761-21723	Sequence 21723, A
4	36.6	9.8	499	10 US-09-783-590-5790	Sequence 5790, Ap
5	36	9.6	381	9 US-10-015-219-700	Sequence 700, App
6	36	9.6	381	10 US-09-777-564-700	Sequence 700, App
7	35.8	9.5	966	9 US-10-213-700-2	Sequence 2, Appl
8	35.8	9.5	204	10 US-09-864-761-21008	Sequence 21008, A
9	35.8	9.5	474	10 US-09-864-761-4255	Sequence 4255, Ap
10	35.4	9.4	876	10 US-09-770-445-544	Sequence 544, App
11	35.4	9.4	1092	9 US-09-938-842A-1125	Sequence 1125, App
12	35	9.3	267	10 US-09-878-574-15076	Sequence 15076, A
13	35	9.3	563	10 US-09-864-761-13293	Sequence 13293, A
14	34.8	9.3	327	10 US-09-864-761-128059	Sequence 128059, A
15	34.8	9.3	1626	10 US-09-764-848-19	Sequence 19, Appl
16	34.8	9.3	5361	9 US-09-742-096-2	Sequence 2, Appl
17	34.8	9.3	6152	9 US-09-742-096-1	Sequence 1, Appl
18	34.6	9.2	438	10 US-09-864-761-4988	Sequence 4988, Ap
19	34.6	9.2	542	10 US-09-864-761-12239	Sequence 12239, A

20	34.6	9.2	8391	9 US-10-151-736-3	Sequence 3, Appl
21	34.6	9.2	8493	9 US-10-151-736-5	Sequence 5, Appl
22	34.4	9.2	846	10 US-09-815-242-8918	Sequence 8918, Ap
23	34.4	9.2	1282	9 US-10-002-344A-89	Sequence 89, Appl
24	34.2	9.1	381	10 US-09-864-761-21610	Sequence 21610, A
25	34.2	9.1	659158	9 US-09-771-208-20	Sequence 20, Appl
26	33.8	9.0	1824	10 US-09-815-242-9281	Sequence 9281, Ap
27	33.8	9.0	1824	10 US-09-815-242-9584	Sequence 9584, Ap
28	33.6	9.0	305	10 US-09-864-761-19262	Sequence 19262, A
29	33.6	9.0	496	10 US-09-864-761-2534	Sequence 2534, Ap
30	33.4	8.9	420	10 US-09-864-761-23266	Sequence 23266, A
31	33.4	8.9	451	10 US-09-864-761-6554	Sequence 6554, Ap
32	33.4	8.9	451	10 US-09-864-761-14152	Sequence 14152, A
33	33.4	8.9	7657	10 US-09-070-927A-33	Sequence 33, Appl
34	33.2	8.9	157	10 US-09-864-761-29856	Sequence 29856, A
35	33.2	8.9	423	9 US-09-938-842A-592	Sequence 592, App
36	33.2	8.9	58985	9 US-09-901-152-3	Sequence 3, Appl
37	33	8.8	1959	10 US-09-864-761-4012	Sequence 4012, Ap
38	33	8.8	2920	10 US-09-801-574-87	Sequence 87, Appl
39	33	8.8	5711	9 US-09-734-672-1	Sequence 1, Appl
40	33	8.8	5711	9 US-09-734-672-3	Sequence 3, Appl
41	33	8.8	5711	9 US-09-734-672-5	Sequence 5, Appl
42	33	8.8	5711	9 US-09-982-828-1	Sequence 1, Appl
43	33	8.8	5711	9 US-09-982-828-3	Sequence 3, Appl
44	33	8.8	5711	9 US-09-982-828-5	Sequence 5, Appl
45	33	8.8	5711	9 US-10-022-819-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-09-864-761-19608/c
Sequence 19608, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DEIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30

Query Match	10.68;	Score 39.6;	DB 10;	Length 766;
Best Local Similarity	51.1%;	Pred. No. 0.051;		
Matches 93; Conservative	0;	Mismatches 89;	Indels 0;	Gaps 0;

QY	187	TACATCGTTTTGAAGACACGGTCAACAGAGAAATGTTACTCATTGAGTCGGTATGGGAA	248
Db	687	TCCACACTTGAAGGACACCATGAGAGATGATGTGTGATGAAGTGAAGATGATGTGTGT	622
QY	247	GAGTTAGACCAAGAAAAATTTGTTGTGCGACGAAAACTTCACGGAATTTATTGTACAGAC	308
Db	627	GATGAGAGGAGATGATTAATTAATGATGACGATGATGATGATTTGATTAATGATGGGAGAG	568
QY	307	TGCGAGGTTAAAGATTCGAGTTAATGCGGACGTCACAGTACGAGTCAGATGAAGTGCAT	366
Db	567	GACGCGAGGAATGATGTGATAGTGCATGATGATGAAAAATGATTAATGAGAGAAATGAGGAT	508
QY	367	GA 368	
Db	507	GA 506	

```

RESULT 2
US-09-864-761-2825/c
; Sequence 2825, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30

```

```

? ORGANISM: Homo sapiens
? FEATURE:
? OTHER INFORMATION: MAP TO AL008720.1
? OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1
? OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.89
? OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 1.1
? OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1
? OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.8
? OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.2
US-09-864 -761-2825

```

Query Match	10.6%;	Score 39.6;	DB 10;	Length 1944;
Best Local Similarity	51.1%;	Pred. No. 0.075;		
Matches 93;	Conservative	0;	Mismatches 89;	Indels 0;
				Gaps 0

QY	187	TACATCGTTTTGAAGACACGCTCACACAGGAATGTTACTACAACTGCTTAATGCTGA	24
Db	924	TCACACCTTGAAGACACACCATGAGAGATGATGGTATGATGACATGAAGATGATGTGT	865
QY	247	GAGTTAGACCAAGAAAAATTTGTTGTGACAGCAAAACTTCACGAAATTAATTTCACAGC	306
Db	864	GATGAGAGAGATGATTAATTAATGATGACAGATGATGATGATTAATGATTAATGATGGAGAG	807
QY	307	TGCGAGGTTAAAGATGCAGGTAAATGCGGACAGGTACAGGTACAGATCACATGAAGTTGAT	366
Db	804	GACGAGAGAAATGATGTGATGATGCGATGATGATGATGATGATGATGATGATGATGATGAT	743
QY	367	GA 368	
Db	744	GA 743	

```

RESULT 3
US-09-864-761-21723
; Sequence 21723, Application US/09864761
; Patent No. US20020048763A1
;
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
;
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1

```

CURRENT APPLICATION NUMBER: US/09/864,761
 CURRENT FILING DATE: 2001-05-23
 PRIOR APPLICATION NUMBER: US 60/180,312
 PRIOR FILING DATE: 2000-02-04
 PRIOR APPLICATION NUMBER: US 60/207,456
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: US 09/632,366
 PRIOR FILING DATE: 2000-08-03
 PRIOR APPLICATION NUMBER: GB 24263,6
 PRIOR FILING DATE: 2000-10-04
 PRIOR APPLICATION NUMBER: US 60/236,359
 PRIOR FILING DATE: 2000-09-27
 PRIOR APPLICATION NUMBER: PCT/US01/00666
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00667
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00664
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00669
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00665
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00668
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00663
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00662
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00661
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00670
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: US 60/234,687
 PRIOR FILING DATE: 2000-09-21
 PRIOR APPLICATION NUMBER: US 09/608,408
 PRIOR FILING DATE: 2000-06-30
 PRIOR APPLICATION NUMBER: US 09/774,203
 PRIOR FILING DATE: 2001-01-29
 NUMBER OF SEQ ID NOS: 49117
 SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
 SEQ ID NO 21723
 LENGTH: 315
 TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE:
 OTHER INFORMATION: MAP TO AL033533.5
 OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.7
 OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.3
 OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.9
 OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.3
 OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.6
 OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.5
 OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.8
 OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.9
 OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.6
 OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.4
 OTHER INFORMATION: NT HIT: U28921.1, EVALUE 6.50e-01
 US-09-864-761-21723

Query Match 9.9%, Score 37.2, DB 10, Length 315;
 Best Local Similarity 50.0%, Pred. No. 0.17;
 Matches 93; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

QY 190 ATCGTTTGAAGACACGCTACACAAAGATGTTACTACATGCGTGGTGAAGAG 249
 DB 52 ACCGAGTTTGAGAAACAAAATAAAGAAAAAATGCCAGACAGAGAACAA 111
 QY 250 TTAGACCAAGAAAAATTTGTCGACGCAAACTTCAGGAAATTTATTGACAGACTGC 309
 DB 112 GAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 171
 QY 310 GAGGTAAAGATGACAGTATGCGGACGATACAGGTACAGACTGATGAAGTTGATCAA 369

DB 172 GAAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 231
 QY 370 GATTAA 375
 DB 232 GAAGAA 237
 RESULT 4
 US-09-783-590-5790
 ; Sequence 5790, Application US/09783590
 ; Patent No. US20020110850A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Dillon, Patrick J.
 ; APPLICANT: Haseltine, William A.
 ; APPLICANT: Li, Haodong
 ; APPLICANT: Rosen, Craig A.
 ; APPLICANT: Ruben, Steven M.
 ; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
 ; FILE REFERENCE: PO-16,2C1
 ; CURRENT APPLICATION NUMBER: US/09/783,590
 ; CURRENT FILING DATE: 2000-02-15
 ; PRIOR APPLICATION NUMBER: 08/420,856
 ; PRIOR FILING DATE: 1995-04-12
 ; PRIOR APPLICATION NUMBER: 08/346,731
 ; PRIOR FILING DATE: 1994-11-21
 ; NUMBER OF SEQ ID NOS: 12485
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 5790
 ; LENGTH: 499
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (9)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: misc feature
 ; LOCATION: (127)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: misc feature
 ; LOCATION: (211)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: misc feature
 ; LOCATION: (233)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: misc feature
 ; LOCATION: (262)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: misc feature
 ; LOCATION: (270)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: misc feature
 ; LOCATION: (302)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: misc feature
 ; LOCATION: (320)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: misc feature
 ; LOCATION: (330)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: misc feature
 ; LOCATION: (337)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: misc feature
 ; LOCATION: (345)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: misc feature
 ; LOCATION: (346)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: misc feature
 ; LOCATION: (348)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: misc feature
 ; LOCATION: (368)

```

OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (375)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (387)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (388)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (391)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (394)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (395)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (396)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (398)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (408)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (417)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (423)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (438)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (439)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (456)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (467)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (468)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (469)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (470)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (476)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (481)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (487)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (492)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (497)
OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-5790
Query Match          9.8%; Score 36.6; DB 10; Length 499;

```

```

Best Local Similarity 50.6%; Pred. No. 0.31;
Matches 87; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 31 CTCGCAAGCTTGCATCTCACTACTCTTCTTCAGAAAGAGTGAAGTTGTGGAGCTTT 90
    || || || || || || || || || || || || || || || || || || || ||
Db 10 CTAGATTCATGTAATCTAAATCTAAACCCCTGCCAGGTAGCCAGATCAATCTTGTGTAC 69

QY 91 TACCGAAGCAAGAAATATACAGACTTGCATTAATCTTTAAGAGTCCTGTAC 150
    || || || || || || || || || || || || || || || || || || || ||
Db 70 TCCAGCAGCAAGAAAGAGATGCTTAGATTAAGACAGTAAAGAAATGTATATCATTAAT 129

QY 151 GAATGCAAAAAAACATGTTCAAGACGAGTACTGCTACATCTTTTGAAG 202
    || || || || || || || || || || || || || || || || || || || ||
Db 130 CATTTCAATGTAATTAATCTTTAATTTGAACATTTAGCAATGCTGTAGATG 181

RESULT 5
US-10-015-219-700
; Sequence 700, Application US/10015219
; Publication No. US20030008299A1
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.493C1
; CURRENT APPLICATION NUMBER: US/10/015,219
; NUMBER OF SEQ ID NOS: 1739
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 700
; LENGTH: 381
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 66, 201, 213, 225, 251, 255, 261, 262, 265, 280, 290, 303,
; LOCATION: 313, 322, 339, 373
; OTHER INFORMATION: n = A,T,C or G
US-10-015-219-700

Query Match          9.6%; Score 36; DB 9; Length 381;
Best Local Similarity 49.5%; Pred. No. 0.42;
Matches 90; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 193 GTTTTGAGACACGCGTCACAAGAAATGTACTACATGCTGTGATGTAAGAGTTA 252
    || || || || || || || || || || || || || || || || || || || ||
Db 21 GTGATGAGAGAGGAGGACGACGAAGAGAGAGATGAGAGACNAGACGATGAGAT 80

QY 253 GACCAAGAAAAATTTGTGTCGACGAAACTTCACGAAATTTATTTGACAGACTGCGAG 312
    || || || || || || || || || || || || || || || || || || || ||
Db 81 GGTGAAGAGAGAGAGTTGATAGAGATGATGAAGATGAAGATGTAAGAGGATGAG 140

QY 313 GGTAAAGATGCAAGTATCGCGCAGGTGACGAGTCAAGTCAAGATGAAGTTGATGAAGAT 372
    || || || || || || || || || || || || || || || || || || || ||
Db 141 GAGCAGAGATGAGTCAAGTGAAGAGAGAAATTTGACCTTGATGAGAGATGATGAAGAT 200

QY 373 TA 374
    |
Db 201 NA 202

RESULT 6
US-09-777-564-700
; Sequence 700, Application US/09777564
; Patent No. US20020022591A1
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.493
; CURRENT APPLICATION NUMBER: US/09/777,564
; CURRENT FILING DATE: 2001-02-05

```

```

RESULT 7
US-10-213-700-2
: Sequence 2, Application US/10213700
: Publication No. US2003002332A1
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
Golil, Surya K.
TITLE OF INVENTION: NOVEL HUMAN PHOSPHATASE INHIBITOR PROTEIN
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/213,700
FILING DATE: 06-Aug-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/766,738
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0177 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX: <Unknown>
INFORMATION FOR SEQ. ID NO.: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 966 base pairs
TYPE: nucleic acid

```

RESULT 8
HS-09-864-761-31008/C

Patent NO. US20020048763A1

GENERAL INFORMATION:

APPLICANT: Penn, Sharon G.

APPLICANT: Rank, David R.

APPLICANT: Hanzel, David K.

APPLICANT: Chen, Wensheng

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY

FILE REFERENCE: Aeomica-X-1

CURRENT APPLICATION NUMBER: US/09/864,761

PRIOR FILING DATE: 2001-05-23

PRIOR APPLICATION NUMBER: US 60/180,312

PRIOR FILING DATE: 2000-02-04

PRIOR APPLICATION NUMBER: US 60/207,456

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: US 09/632,366

PRIOR FILING DATE: 2000-08-03

PRIOR APPLICATION NUMBER: GB 24263.6

PRIOR FILING DATE: 2000-10-04

PRIOR APPLICATION NUMBER: US 60/236,359

PRIOR FILING DATE: 2000-09-27

PRIOR APPLICATION NUMBER: PCT/US01/00666

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00667

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00664

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00669

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00665

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00668

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00663

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00662

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00661

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00670

PRIOR FILING DATE: 2001-01-30


```

; APPLICANT: Lediord, Brooke L.
; APPLICANT: Woessner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Krieker, Maja
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hurban, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; FILE REFERENCE: 202305 (PARA-012PRV)
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: US/09/770,445
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 999
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 544
; LENGTH: 876
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-770-445-544

```

```

Query Match
Best Local Similarity 9.4%; Score 35.4; DB 10; Length 876;
Matches 78; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

```

```

QY 227 ACAATGTCGTTGATGAGTGAAGTGAACCAAGAAATTTGTCGACGAAACTTCA 286
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 410 ACATGATCTCGAAGCAGCTGATCGATGAGATGAGTGTGGAGCTGCAAGATTTA 351
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 287 CGGAAATTTATTTGACAGCTCGAGGCTGAAGATGCGAGTATGCGGACGATCAGGTG 346
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 350 TTGTGATTAACCTGAAGAAATGGTGAAGATTGATGAGAAAGATATTCTGGGATATG 291
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 347 ACAGTCAGATGAAGTTGATGAAGATTAA 375
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 290 ATGAGTTCGAGCACTTGATTACGATGAA 262
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

```

RESULT 11
US-09-938-842A-1125
; Sequence 1125, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Krieps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCRI1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 1125
; LENGTH: 1092
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-1125

```

```

Query Match
Best Local Similarity 9.4%; Score 35.4; DB 9; Length 1092;
Matches 78; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

```

```

QY 227 ACAATGTCGTTGATGAGTGAAGTGAACCAAGAAATTTGTCGACGAAACTTCA 286
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 881 ACATGATCTCGAAGCAGCTGATCGATGAGATGAGTGTGGAGCTGCAAGATTTA 940
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 287 CGGAAATTTATTTGACAGCTCGAGGCTGAAGATGCGAGTATGCGGACGATCAGGTG 346
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 941 TTGTGATTAACCTGAAGAAATGGTGAAGATTGATGAGAAAGATATTCTGGGATATG 1000
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 347 ACAGTCAGATGAAGTTGATGAAGATTAA 375
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1001 ATGAGTTCGAGCACTTGATTACGATGAA 1029
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

```

RESULT 12
US-09-878-574-15076
; Sequence 15076, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 15076
; LENGTH: 267
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: 701069621H1
US-09-878-574-15076

```

```

Query Match
Best Local Similarity 9.3%; Score 35; DB 10; Length 267;
Matches 59; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

```

```

QY 277 GAAACTTCACGGAATTTTGAACAGCTCGAGGCTAAAGATGCAAGTATCGCGCA 336
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 54 GAGAAGTCACCTGACCTTATTCATTAGAGAGGTGATGAGTGAAGCTGATGATGAG 113
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 337 GGTACAGGTGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 375
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 114 GGAAGGAGGATGAATCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 152
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

```

RESULT 13
US-09-864-761-13293
; Sequence 13293, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Weosheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeonica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27

```

```

PRIORITY APPLICATION NUMBER: PCT/US01/00666
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00667
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00664
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00669
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00665
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00668
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00663
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00662
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00661
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00670
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: US 60/234,687
PRIORITY FILING DATE: 2000-09-21
PRIORITY APPLICATION NUMBER: US 09/608,408
PRIORITY FILING DATE: 2000-06-30
PRIORITY APPLICATION NUMBER: US 09/774,203
PRIORITY FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 13293
LENGTH: 563
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC012404.4
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.8
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.8
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.2
US-09-864-761-13293

Query Match          9.3%: Score 35; DB 10; Length 563;
Best Local Similarity 46.5%: Pred. No. 0.95;
Matches 113; Conservative 0; Mismatches 130; Indels 0; Gaps 0;

QY 133 AAGTCTGTGATCTTGACGATGCAAAAAACATGTTTCAAGACGGAGTACTGCTACATC 192
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 71 AATCACTCTAATCAAGAAAAAGGATTTGTTTCTTCAATTAATTAATTAATTTCTTAAA 130

QY 193 GTTTTGAAGACACGCTCAACAAAGAAATGTTTACTACATGTCGTTGATGTAAGAGTTA 252
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 131 TGTAAAGATGAAAAAATGTAAGATGATTAATAAACAGTTCCACAGAAAGAAAGAA 190

QY 253 GACCAAGAAAAATTTGTTTGCAGAAAACTTCAGGAAAAATTTATTGACAGACTGGAG 312
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 191 GAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 250

QY 313 GGTAAAGATGACAGTATGCGGACAGTACAGGATGACGAGTACAGTGAAGTATGATGAT 372
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 251 GAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 310

QY 373 TAA 375
    ||
Db 311 GAA 313

```

RESULT 14
 US-09-864-761-28059/c
 Sequence 28059, Application US/09864761
 Patient No. US20020048763A1
 GENERAL INFORMATION:
 APPLICANT: Penn, Sharon G.
 APPLICANT: Rank, David R.
 APPLICANT: Hanzel, David K.

```

APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
FILE REFERENCE: Aecm1ca-x-1
CURRENT FILING DATE: 2001-05-23
PRIORITY APPLICATION NUMBER: US 60/180,312
PRIORITY FILING DATE: 2000-02-04
PRIORITY APPLICATION NUMBER: US 60/207,456
PRIORITY FILING DATE: 2000-05-26
PRIORITY APPLICATION NUMBER: US 09/632,366
PRIORITY FILING DATE: 2000-08-03
PRIORITY APPLICATION NUMBER: GB 24263.6
PRIORITY FILING DATE: 2000-10-04
PRIORITY APPLICATION NUMBER: US 60/236,359
PRIORITY FILING DATE: 2000-09-27
PRIORITY APPLICATION NUMBER: PCT/US01/00666
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00667
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00664
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00669
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00665
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00668
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00663
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00662
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00661
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00670
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: US 60/234,687
PRIORITY FILING DATE: 2000-09-21
PRIORITY APPLICATION NUMBER: US 09/608,408
PRIORITY FILING DATE: 2000-06-30
PRIORITY APPLICATION NUMBER: US 09/774,203
PRIORITY FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 28059
LENGTH: 327
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC005668.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.55
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5
US-09-864-761-28059

Query Match          9.3%: Score 34.8; DB 10; Length 327;
Best Local Similarity 51.3%: Pred. No. 0.87;
Matches 81; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 217 GAATGTACTACATGTCGTTGATGTAAGAGTTAGACCAAGAAAAATTTGTTGAC 276
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 320 GATGATGACAGAGTATGATGATGATGATGACAGCAGTATGATGATGATGATGAT 261

QY 277 GAAACTTCAGCGAAAAATTTATTGACAGACTGCGAGGAGTAAGATGACAGTATGCGCA 336
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 260 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 201

QY 337 GGTACAGGTGACGAGTACAGTGAAGTGAAGATTA 374
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 200 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 163

```


RESULT 15

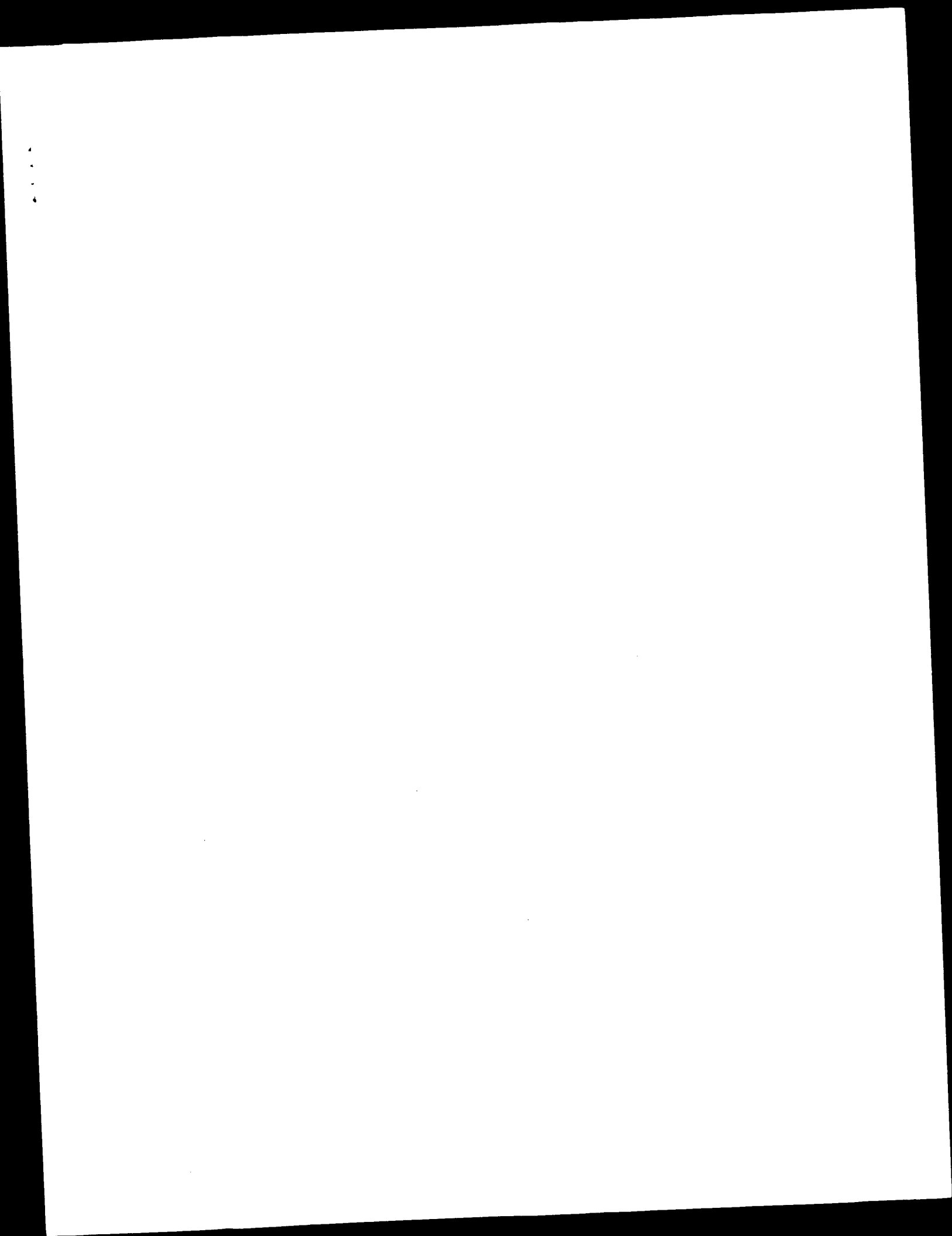
US-09-764-848-19
; Sequence 19, Application US/09764848
; Patent No. US20020077270A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT208
; CURRENT APPLICATION NUMBER: US/09/764,848
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 1626
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-848-19

Query Match

Best Local Similarity 9.3%; Score 34.8; DB 10; Length 1626;
Matches 78; Conservative 1; Mismatches 73; Indels 0; Gaps 0;

QY	217	GAATGTTACTACAAATGTCGTGATGCTGAGAGTTAGACCAAGAAATTTGTTGTCGAC	276
DB	459	GCATCTATATTAAGAGAGTTGAAGATTAAGATCAAGAGGAGAGAGAGATGAGAT	518
QY	277	GAATCTCAGCGAAATTTATTTGACAGACTGCGAGSGGTAAAGATGCAGTAAATGGGCA	336
DB	519	GAAGATCTTTCTAAATATTAAGTAAATGAGATGAGGATGAGGATGAAGATGACGCTGATCTCA	578
QY	337	GGTACAGGTGAGAGCTGAGATGAGTGAATGATGA	610
DB	579	AAATATTAATCTTGATGCCAGTGAAGAGAGAGA	610

Search completed: February 25, 2003, 02:24:22
Job time : 65 secs



GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 25, 2003, 02:23:10 : Search time 22 Seconds
(without alignments)
3277.310 Million cell updates/sec

Title: US-09-936-737A-1
Perfect score: 674
Sequence: 1 atgaagatattctgtattc.....atgaagtgtatgaagattaa 375

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 283224 segs, 96134422 residues

Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+_n2p.model -DEV=xlh
-Q/cgn2_1/USPTO_spool/US0936737.runal_14022003_100558_1407/app-query.fasta_1.519
-DB=PIR_73 -OFMT=fastan -SUFFIX=prp -MINMATCH=0.1 -LOOPEXT=0 -LIST=45
-UNITS=bits -START=1 -END=1 -MATRIX=blomsun62 -TRANS=humana4.0.cdi -LIST=45
-DOCALLIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MTN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=prp -NORM=ext -HEAPSIZE=500 -MTNLEN=0 -MAXLEN=200000000
-USER=US0936737/cgn2_1_16.ernal_14022003_100558_1407 -NCP=6 -ICPU=3
-NO_XLPHY -NO_MAP -LARGEQUERY -NEG.SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-NARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: PIR_73:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	88.5	13.1	147	A42435	leech antiplatelet
2	84.5	12.5	164	T33824	hypothetical prote
3	82.5	12.2	412	A55320	immunophilin FKBP4
4	82	12.6	637	T00548	hypothetical prote
5	82	12.2	907	T27317	hypothetical prote
6	80.5	11.9	1303	B84493	probable Athila re
7	79.5	12.2	507	S64507	bone morphogenetic
8	79.5	11.8	1038	JC5527	PAR interacting pr
9	79.5	11.8	1277	T37331	hypothetical prote
10	78.5	11.6	264	T23054	two-component sens
11	78.5	11.6	550	T26562	hypothetical prote
12	78	11.6	682	B84021	hypothetical prote
13	78	11.6	797	A36811	hypothetical prote
14	77.5	11.5	191	T15129	hypothetical prote

15	77	11.4	555	2	G90594	hypothetical prote
16	76.5	11.4	168	2	A96554	hypothetical prote
17	76.5	11.4	786	2	C86406	88.6k hypothetical
18	76	11.3	559	2	A53423	tpcct-gamma protel
19	76	11.3	1255	2	B97104	hypothetical prote
20	76	11.3	1278	2	B70236	hypothetical prote
21	75.5	11.2	336	2	T08538	hypothetical prote
22	75.5	11.2	491	2	T16354	hypothetical prote
23	75	11.5	344	2	T40167	hypothetical prote
24	75	11.1	360	2	T34510	hypothetical prote
25	75	11.5	534	2	T39903	serine-rich protel
26	75	11.1	2410	2	T43731	cell wall alpha-g1
27	75	11.1	3844	2	T18402	asparagine/asparta
28	74.5	11.1	146	1	PSDG	phospholipase A2 (
29	74.5	11.1	146	1	PSPGA	phospholipase A2 (
30	74.5	11.1	328	2	S67570	hypothetical prote
31	74.5	11.1	643	2	T21428	hypothetical prote
32	74	11.0	296	2	D70465	hypothetical prote
33	74	11.0	600	2	S07638	spore coat protein
34	74	11.0	612	2	B84809	hypothetical prote
35	74	11.0	4466	1	S17231	dynein beta heavy
36	73.5	10.9	208	2	T06492	ribosomal protein
37	73.5	11.3	222	2	E81564	conserved hypotet
38	73.5	11.3	228	2	H72103	ypdp hypothetical
39	73.5	11.3	228	2	A86518	hypothetical prote
40	73.5	10.9	357	2	T34012	hypothetical prote
41	73.5	10.9	562	2	B41035	chitinase (EC 3.2.
42	73.5	10.9	562	2	S50371	chitinase (EC 3.2.
43	73.5	11.3	636	2	S63131	probable membrane
44	73.5	10.9	2339	2	A45597	DNA-directed RNA p
45	73.5	10.9	4572	2	S57908	hypothetical 527k

ALIGNMENTS

RESULT 1
A42435
leech antiplatelet protein precursor - Mexican leech
C:Species: Haementeria officinalis (Mexican leech)
C:Date: 07-Apr-1994 #sequence: revision 07-Apr-1994 #text: change 07-May-1999
C:Accession: A42435; A42434
R:Keller, P.M.; Schultz, L.D.; Condra, C.; Karczewski, J.; Connolly, T.M.
J. Biol. Chem. 267, 6899-6904, 1992
A:Title: An inhibitor of collagen-stimulated platelet activation from the salivary gl
A:Reference number: A42435; MUID:92202247; PMID:1551898
A:Accession: A42435
A:Molecule type: mRNA
A:Residues: 1-147 <KE>
A:Cross-references: GB:M81489
A:Note: the authors translated the codon AGC for residue 20 as Arg and GAA for residu
A:Note: the authors' translation is shown at residue 65
R:Connolly, T.M.; Jacobs, J.W.; Condra, C.
J. Biol. Chem. 267, 6893-6898, 1992
A:Title: An inhibitor of collagen-stimulated platelet activation from the salivary gl
A:Reference number: A42434; MUID:92202246; PMID:1551897
A:Accession: A42434
A:Molecule type: protein
A:Residues: 59-91;101-113;123-139 <CON>
F:1-21/Domain: signal sequence #status predicted <SIG>
F:122-147/Product: leech antiplatelet protein #status predicted <MAT>
Alignment Scores:
Pred. No.: 0.366
Score: 88.50
Percent Similarity: 35.25%
Best Local Similarity: 24.59%
Query Match: 13.13%
Length: 147
Matches: 30
Conservative: 13
Mismatch: 32
Indels: 47
Gaps: 4
US-09-936-737A-1 (1-375) x A42435 (1-147)
QY 1 ATGAAGTATTCTGATTTCTCTTCCTTCCTGCGCAAGCTTGATGATCTCA----- 51

```

Db 1 MetasensPheLeuPheSerLeuAlaLys-----SerLeuValAlaIleProAla 18
OY 51 -----
Db 19 IleSerAlaGlnAspGlnAspAlaGlyValaGlyAspGluThrSerGluGlyGluAsp 38
OY 52 ACTACTCTTCAGAGAA-----
Db 39 ThrThrGlySerAspGluThrProSerThrGlyGlyGlyAspGlyValAsnGluGlu 58
OY 70 -----CGTGAAGATTGTTGACGTTTACGCCAACAAGAAATATACAGAC 114
Db 59 ThrIleThrAlaGlyAsnGlyAspCysThrSerLysArgProGlyThrLysLeuProAsp 78
OY 115 TTCGATTAATCTTTAAGACGCTTCATCTTGACGCAATCCAAAAACATGTTTACG 174
Db 79 AsnLeuLeuThrLysThrGluPheThrSerValAspGlyLysArgGlyMetCysGluGlu 98
OY 175 ACGGATAC-----TGCTACATCGTTTGAAGACAGCGTCAACAGCAAGATGTTAC 225
Db 99 SerAlaValaGluProSerCysThrLysLeuGlnIleAsnThrGluThrAsnGluCysThr 118
OY 226 TACAAAT 231
Db 119 ArgAsn 120

```

RESULT 2

T32824
hypothetical protein F54D12.1 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*

C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

R:Magdi, L.; Goela, D.

submitted to the EMBL Data Library, December 1997

A:Description: The sequence of C. elegans cosmid F54D12.

A:Reference number: Z21228

A:Accession: T32824

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-164 <MAG>

A:Cross-references: EMBL:AF040647; PIDN:AA94992.1; GSPDB:GN00020; CESP:F54D12.1

A:Experimental source: strain Bristol N2; clone F54D12

A:Genetics:

A:Gene: CESP:F54D12.1

A:Map position: 2

A:Introns: 91/3; 138/1

Alignment Scores:

```

Pred. No.: 0.975      Length: 164
Score: 84.50      Matches: 23
Percent Similarity: 45.35%      Conservative: 16
Best Local Similarity: 26.74%      Mismatches: 34
Query Match: 12.54%      Indels: 13
Db: 2      Gaps: 3

```

US-09-936-737A-1 (1-375) x T32824 (1-164)

```

OY 10 TTCTGANTTCCTTCCTTCGCTGCAAGCTGCTGATCTCACTACTCTTCAGAGAA 69
Db 5 PheLeuIleCysThrValLeuValAsnSerIleMetValSerValaArgValaProGlu 24
OY 70 CGTGAAGATTGTTGACGTTTACGCCAACAAGAAATATACAGCTTCGATTAATCTTT 129
Db 25 IleSer-----ThrSerThrSerThrGluIle 33
OY 130 AAGAAGTCCTCTGATCTTGACGAATGCAAAAAACATGTTTCAAGACGAGTACTGCTAC 189
Db 34 IleAsnSerLeuThrTrpAspGluCysValIleGlnCysLeuSerThrGluIleCysIle 53
OY 190 ATCGTTTTTGAAGACAGCGTCAACAAGCAATGTTAC---TACATGTCGTTGATGGTGA 246
Db 54 MetAlaTrpSerAsnSerLeuAsn---IleCysThrLeuThrAlaValaGlyAspValIle 72

```

```

OY 247 GAGTAGACCAAGAAAA 264
Db 73 GluValaArgHisAspGln 78

```

RESULT 3

A55320

immunophilin FKBP46 - fall armyworm

C:Species: *Spodoptera frugiperda* (fall armyworm)

C:Date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 02-Sep-2000

C:Accession: A55320

R:Alnemri, E.S.; Fernandes-Alnemri, T.; Pomeroy, K.; Robertson, N.M.; Dudley, K.; D

J. Biol. Chem. 269, 30828-30834, 1994

A:Reference number: A55320; MID:95074110; PMID:7527037

A:Accession: A55320

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-412 <ALN>

A:Cross-references: GB:U15038; NID:9595844; PID:9595845

C:Superfamily: Yeast peptidylprolyl isomerase FR3; BKB-type peptidylprolyl isomerases

C:Keywords: nucleus; phosphoprotein

F:324-371/Domain: BKB-type peptidylprolyl isomerase homology <PPI>

Alignment Scores:

```

Pred. No.: 1.67      Length: 412
Score: 82.50      Matches: 32
Percent Similarity: 41.74%      Conservative: 16
Best Local Similarity: 27.83%      Mismatches: 40
Query Match: 12.24%      Indels: 27
Db: 2      Gaps: 5

```

US-09-936-737A-1 (1-375) x A55320 (1-412)

```

OY 43 CTGATCTCAACTACTCTTCACAGACGTAAGATTGTCAGCTTTACCGCAACAGA 102
Db 116 LeuValProAlaLysAsnLysArgLysLeuGluAsnAlaAsnAspAlaThrAsnLys 135
OY 103 AATATACAGACTTCATTAATCTTTTACAGCTC-----TCGATCTT 147
Db 136 LysAlaLysProAspLysLysAlaGlyLysAsnSerAlaProAlaAlaLysSerAsp 155
OY 148 GAGCAATGCAAAAAACATGTTTCAAGACGAGTACTCTACATCGTTTGAAGACAG 207
Db 156 AspAspAspAsp-----GluAspGln 162
OY 208 GTCAACAGCAATGTTTACTACAAATGCTTGATGCTGAAGATTGACCAAGAAATTT 267
Db 163 LeuGlnLys-----PheLeuAspGlyGluAspIleAspThrAspGlu--- 176
OY 268 GTTGTGACGAAACCTTCACGAAATTTTTCACAGACTCGGAGGTTAAAGATGACAGT 327
Db 177 ---AsnAspGluSerPheLysMetAsn-----ThrSerAlaGluGlyAspAspSerAsp 193
OY 328 AATCGCGCAGTACAGAGTGCAGTCAGATGAGTGAAGATTGAAGAT 372
Db 194 GluGluAspAspAspGluAspGluGluAspGluGluAspAsp 208

```

RESULT 4

T00548

hypothetical protein At2g39380 (imported) - *Arabidopsis thaliana*

N:Alternate names: hypothetical protein F12L6.4

C:Species: *Arabidopsis thaliana* (mouse-ear cress)

C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 16-Feb-2001

C:Accession: T00548; F84816

R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; K

submitted to the EMBL Data Library, July 1998

A:Description: *Arabidopsis thaliana* chromosome II BAC F12L6 genomic sequence.

A:Reference number: Z14168

A:Accession: T00548

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-637 <ROU>

A:Cross-references: EMBL:AC004218; NID:g3355463; PID:g3355467
A:Experimental source: cultivar Columbia
R.Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Feb-2001 #sequence_revision 02-feb-2001 #text_change 02-Feb-2001
C:Accession: B84493
R.Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
euss, D.; Niernan, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
Nature 402, 761-768, 1999
A>Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: F84816
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-637 <STO>
A:Cross-references: GB:AEO02093; NID:g3355467; PIDN:AAC37829.1; GSPDB:GN00139
A:Gene: FL2L6.4; Atg39380
A:Map position: 2
C:Superfamily: tomato leucine zipper-containing protein

Alignment Scores:
Pred. No.: 1.93 Length: 637
Score: 82.00 Matches: 38
Percent Similarity: 48.91% Conservative: 29
Best Local Similarity: 27.74% Mismatches: 30
Query Match: 12.60% Indels: 40
Gaps: 9

US-09-936-737A-1 (1-375) x T00548 (1-637)

OY	365	TCAACTTCATCTGCAGCGTCA-----CCTGTACCTGCCGGCATTAACCTGCA	321
Db	11	SerThrsrSerProlyserPheleuseSerPheproSerPheThrSerLeuproAla	30
OY	320	TCTTTACC-----TCGACGTCTGCAAAATTTCGCGTAGAGTTTG	276
Db	31	Ser--ProleusenglnThrpheserGlnSerMetcglunThrValglnAlala	49
OY	275	TCGACACAATTTCCTTGGCTACCTTCACATCACAGACATTTGATACATCC	216
		:::	
Db	50	GluSerlleIlellylSylTrpAspProasnsrProser-----	62
OY	215	TTGTGACCGTCTCTCAAAAACGATGTACAGTACTCCGCTCTGAACATTTTTTTG	156
Db	63	-----TyrThrlysllellesearlepheser	71
OY	155	CATTCGTCAAGA---TCAGAGACTTC-----TTAAAGATTTCAGAGTCTGTA	108
Db	72	HlsSrAalgauglualielysglurhelleargyslleargspneulrgralawet	91
OY	107	TATTTTCCTTTCGCGCTAAACGTCACATCTTCACAGTCTCTCTGAAGAGTAGTT---	51
		:::	
Db	92	HlsPhelau-----lleSerGlhlssrlysrSerAlaylsleuVallleuala	107
OY	50	---GAGATCAGCAGCTTGCGGCGCAAGAGAAATCAAGAAATCAACTTC	3
		::: :::::	
Db	108	GlhlHlsleuMetsclInlleAlametaArgleuGu---lysGlurPhephe	123

RESULT 5
T27317
hypothetical protein Y69H2.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T27317
R:McMurray, A.
submitted to the EMBL Data Library, August 1997
A:Reference number: Z20343
A:Accession: T27317
A:Molecule type: DNA
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Residues: 1-907 <WTL>
A:Cross-references: EMBL:Z98877; PIDN:CAB1570.1; GSPDB:GN00023; CESP:Y69H2.2
A:Experimental source: clone Y69H2
C:Genetics:
A:Gene: CESP:Y69H2.2

Alignment Scores:
Pred. No.: 1.97 Length: 907
Score: 82.00 Matches: 37
Percent Similarity: 39.85% Conservative: 16
Best Local Similarity: 27.82% Mismatches: 40
Query Match: 12.17% Indels: 40
Gaps: 9

US-09-936-737A-1 (1-375) x T27317 (1-907)

OY	46	ATCTCACTACTTCTTCAGAGAACGTGACAGATTGTCAGCTTT-----	90
Db	206	IleGlyThrHisGlthrlystlyrlyspaspaplysmethrseraspysCysAnglu	225
OY	91	TACCGCAGAGAAAAATATACAGCTTCGATTAATCTTTAAAG-----	132
Db	226	TyralaglInglnlyserThrsp---AspystrlleuTyrlleuthleuCysgly	244
OY	133	-----AAGTCCTGATCTT-----	153
Db	245	GlutrpyskmetvalserglinspwetheuVallysaspsavalsermethylsAspasp	264
OY	154	TCGAAAAAACATGT-----TTCAAAGCAGAGTACTCTACATCGTTTGAAGACACG	207
Db	265	CyslysyserlysglylglyspsserserglytryCys-----Glylys	279
OY	208	GTCACACAGAGATGTACTACAAATGCTGTGATGCGAGAGTTGACACAGAAAAATT	267
Db	280	ValnsnserArgystyle-----ValtyrglnAlaglyGlmeGlu-----	293
OY	268	GTTGTCGACGAAACTTCACGAAATTTATTTGACAGCTCGGAGGTAAAGATGACAGT	327
Db	294	--ThraspProasnAlacysgluspaspaserThrleuCys---GlyAlaglnleuGly	311
OY	328	AATGCGCAGGTACAGGTGACAGTCAAGATGAGTTGAT	366
		:::	
Db	312	HlsGlymetCyslleasntrpGlnseraspValThrAsp	324

RESULT 6
B84493
probable Athlia retroelement ORF1 protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Feb-2001 #sequence_revision 02-feb-2001 #text_change 02-Feb-2001
C:Accession: B84493
R.Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
euss, D.; Niernan, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
Nature 402, 761-768, 1999
A>Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: B84493
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1303 <STO>
A:Cross-references: GB:AEO02093; NID:g4263544; PIDN:AAD15358.1; GSPDB:GN00139
A:Gene: AT2G10660
A:Map position: 2

Alignment Scores:
Pred. No.: 2.89 Length: 1303
Score: 80.50 Matches: 31
Percent Similarity: 42.74% Conservative: 19
Best Local Similarity: 26.50% Mismatches: 38
Query Match: 11.94% Indels: 29
Gaps: 6

US-09-936-737A-1 (1-375) x B84493 (1-1303)

OY	40	TTGCTGATCTCAACTACTTCT-----TCAGAGAACGTGAGAGATTGTTGG	84
----	----	----------------------------------------------------	----

```

Db      166 MetLeuLeuAspThrThrSerAsnGlyAsnPhleuAsnLysHisValGluGluGlyTyr 185
QY      85 ACSTTTTACGACGACGAAATATACA-----GACTTCGATTAATCT 126
Db      186 GluLeuValGluAsnLeuAlaGlnSerAspLysAsnGlyAsnLysAspTyrGluArgSer 205
QY      127 TTTAAGAAAGTCCTCGATCTTGACGAATGCAAAAACATGT-----TTCAAGACGGAG 180
Db      206 IleArgThrSerSerAspSerAspGlu-----LysHisCysArgGluMetLysLysHis 223
QY      181 TACTCTACATCGTTTGTGAAGACACGGCTCAACAAGCAATGTACTACATGTCGTGAT 240
Db      224 ValHisPheLeuGlyAspAspGluThr-----PheGlnValGlnAsp 237
QY      241 GGTGAAGATTGACCAAGAAAATTTGTTGTGACGAAACTCTCAGGAAATATTATTTG 300
Db      238 GlyGluThrLeuGlnSerGluLysVal-----AsnTyrVal 249
QY      301 ACAGCTGCGAGGCTAAAGATGCAAGTAAATCGCGGACGATACAGTGACGAG 351
Db      250 GlnAsnHisGlyTyrAsnLysGlyLysSerIleGlnThrProLysGlu 266

```

RESULT 7

probable membrane protein YGR189c - yeast (Saccharomyces cerevisiae)
 N:Alternate names: hypothetical protein G7553
 C:Species: Saccharomyces cerevisiae
 C>Date: 17-May-1996 #sequence_revision 17-May-1996 #text_change 19-Apr-2002
 R:Arroyo, J.; Garcia-Gonzalez, M.; Garcia-Saez, M.I.; Sanchez-Perez, M.; Nombela, C.
 submitted to the Protein Sequence Database, May 1996
 A:Reference number: S64499
 A:Accession: S64507
 A:Molecule type: DNA
 A:Residues: 1-507 <ARR>
 A:Cross-references: EMBL:272974; NID:g1323335; PID:e243566; PID:g1323336; GSPDB:GN00007;
 A:Experimental source: strain S288C
 C:Genetics:
 A:Gene: SGD:CRM1; MIPS:YGR189C
 A:Cross-references: SGD:S0003421
 A:Map position: 7R
 C:Keywords: transmembrane protein
 F:6-22/Domain: transmembrane #status predicted <TM>

Alignment Scores:

Pred. No.:	Length:	Score:	Matches:	Conservative:	Best Local Similarity:	Mismatches:	Indels:	Gaps:
3.5	507	79.50	42	15	31.82%	50	25	4
43.18%								
12.21%								

US-09-936-737a-1 (1-375) x S64507 (1-507)

```

QY      371 TCTTCATCAACT-----TCATCTGACTGCTACCTGACCTGCCGCACTTACCTGACATCT 318
Db      308 SerSerSerThrValSerSerSerSerThrValSerSerSerValSerSerThr 327
QY      317 TTACCCCGACGCTGCTCAATAATTTCCGTGAAGTTTGTGACACAAATAATTTTCT 258
Db      328 ValSerSerSerValSerSerSerThrValSerSerSerValSerSerThrValSerSer 347
QY      257 TGTCTTAAGTCTTTCACATCAAGACATTTAGTAACCTTCTGTTGACCGGTCTTCA 198
Db      348 SerSerValSerSerSerSerSerThrSerProSerSerSerThrAlaThr-----SerSer 366
QY      197 AAACAGATAGTACACTCCGCTTG----- 171
Db      367 LysThrLeuAlaSerSerSerValThrThrSerSerSerIleSerSerPheGluLysGln 386
QY      170 -----AAACATGTTTCTTTCGATTCGTCAGACATCAAGACATCTTTTAAA 126

```

```

Db      387 SerSerSerSerSerLysThrValAlaSerSerSerThrSerGluSerIleSer 406
QY      125 GAT-----TTATGAGATGCTGATATTTTCTGTCGCGTAAACGTC 84
Db      407 SerThrLysThrProAlaThrValSerSerSerThrThrArgSerThrValAlaProThrThr 426
QY      83 CACCAATCTTCACGCTTCTTCTCAAGAACTAGTGG 48
Db      427 GlnGlnSerSerValSerSerSerProValGln 438

```

RESULT 8

JC5527
 bone morphogenetic protein type II receptor precursor - mouse
 C:Species: Mus musculus (house mouse)
 C>Date: 02-Sep-1997 #sequence_revision 05-Sep-1997 #text_change 24-Sep-1999
 C:Accession: JC5527
 R:Beppu, H.; Minowa, O.; Miyazono, K.; Kawabata, M.
 Blochem. Biophys. Res. Commun. 235, 499-504, 1997
 A:Title: cDNA cloning and genomic organization of the mouse BMP type II receptor.
 A:Reference number: JC5527; MUID:97350808; PMID:9207184
 A:Accession: JC5527
 A:Molecule type: mRNA
 A:Residues: 1-1038 <REP>
 A:Cross-references: DDBJ:AF003942; NID:92253704; PIDN:AA63042.1; PID:92253705
 C:Comment: This protein is a serine/threonine kinase receptor that forms a heteromeric
 C:Genetics:
 A:Gene: BMPR-II
 C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hom
 F:1-26/Domain: signal sequence #status predicted <SIG>
 F:151-174/Domain: transmembrane #status predicted <TM>
 F:201-508/Domain: protein kinase homology <KIN>
 F:202-500/Domain: serine/threonine kinase #status predicted <STK>

Alignment Scores:

Pred. No.:	Length:	Score:	Matches:	Conservative:	Best Local Similarity:	Mismatches:	Indels:	Gaps:
3.65	1038	79.50	27	26	40.46%	35	43	4
40.46%								
20.61%								
11.80%								

US-09-936-737a-1 (1-375) x JC5527 (1-1038)

```

QY      10 TTTCTGATTTCTCTCTGCTGCGCAAGCTTGCTCAATCTTCTTCAAGAA 69
Db      9 PheArgValProTyrPheLeuThrPheValLeuLeuValSerThrThrAlaAlaSerGln 28
QY      70 CGTGAAGATTTGTGACGCTTTTACGCGACGAAATAATATACAGACTTGATTAATCTTT 129
Db      29 AsnGlnGln-----ArgLeuCysAlaPheLysAspProTyr 40
QY      130 AAGAGTCTCTGTGATCTTGACGAATGCAAAAACA-----TGT 168
Db      41 GlnGlnAspLeuGlyIleGlyGlnSerArgIleSerHisGluAsnGlyThrIleLeuLys 60
QY      169 TTCAAGACGAGTACTCTACATCGTTTGTGA-----GACACGTC 210
Db      61 SerLysGlySerThrCysTyrGlyLeuTyrGlyLysSerLysGlyAspIleAsnLeuVal 80
QY      211 AACAGGAATGTTACTACATGTCGTTGATGTGAAGAGTTAGACCAAGAAAATTTGTT 270
Db      81 LysGlnGlySerThrSerHisIleGlyAspProGlnGlyCysHisTyrGluGluCysVal 100
QY      271 GTC----- 273
Db      101 ValThrThrThrProProSerIleGlnAsnGlyThrTyrArgPheCysCysSerThr 120
QY      274 -----GACCAAACTTCACGAGAAAATAT 297
Db      121 AspLeuCysAsnValAsnPheThrGlnAsnPhe 131

```

RESULT 9

T32731

Score: 78.50
Percent Similarity: 41.54%
Conservative: 19

136 TCC---TCTGATCTTGACGAATGCAAAAACATGTTTCAAGACGAGTACTGCTACATC

ATC 192

Db 210 AsnLeuThrSerSerAspGluCysValLysTyrCysTrpLysAspLeuAsnCysPheVal 229
 QY 193 GTTTTGAACACGCGTCAACACAGAAATGTTACTACAAATGCGTTGATGTTGAAGATTA 252
 Db 230 AsnTyrTrpAspLysAspSerLysGluCysTrpTrpPhePheIleAspAsnValHisPhe 249
 QY 253 GACCAAGAAAAATTTGTTGTCGACGAAAC 282
 Db 250 ---LeuGluLysValHisProSerGluAsn 258

RESULT 12

B84021

two-component sensor histidine kinase involved in chemotaxis cheA [Imported] - Bacillus

C:Species: Bacillus halodurans

C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001

C:Accession: B84021

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hirata

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and

A:Reference number: A83650; MUID:20512582; PMID:11058132

A:Accession: B84021

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-682 <STO>

A:Cross-references: GB:AP001517; GB:BA000004; NID:G10175500; PIDN:BA06689.1; GSPDB:GNOC

A:Experimental source: strain C-125

C:Genetics:

A:Gene: cheA

C:Superfamily: chemotaxis protein cheA

Alignment Scores:

Pred. No.:	Length:	Score:	Matches:	Conservative:	Best Local Similarity:	Mismatches:	Indels:	Query Match:
5.13	682	78.00	28	13	47.67%	31	14	11.57%
2	5	11.57%	14	14				

US-09-936-737a-1 (1-375) x B84021 (1-682)

QY 91 TACGGCAAGAAATATACAGACTTGCATAATCTTTAAGAATGCTCTGAT----- 144
 Db 156 TyrAspGluPheGluMetThrValLeuAspGlnSerPheGluGlnGlyTyrGlnAlaTyr 175
 QY 145 -----CTTGACGAATGCCAAAAAACAATGTTTCAAGACGGAGTACTGCTAC 189
 Db 176 GlnIleGluValThrLeuAspGlu-----LysThrLeuLeuLysAlaAlaTyrValPhe 193
 QY 190 ATCGTTTGTGA-----GACACGGTCAACACAGAAATGTTACTACAAATGCTGTGATGT 243
 Db 194 MetValPheGluValLeuGluGlnValGly---GluValIleLysSerThrProSerAla 212
 QY 244 GAGAGCTTACCAACAAAAATTTGTTGTCGACGAAACCTTCACGCAAAATATATTGACA 303
 Db 213 GlnGluLeuGlnGluGlnLysPhe-----AspGluArgPheLeuValThrLeuLeuThr 230
 QY 304 GACTGGCAGAGGTAAAGAT 321
 Db 231 LysValAspGlyGluGlu 236

RESULT 13

A36811

hypothetical protein ORF48 - saimiriine herpesvirus 1 (strain 11)

C:Species: saimiriine herpesvirus 1

A:Note: host Saimiri sciureus (Common squirrel monkey)

C>Date: 16-Oct-1992 #sequence_revision 16-Oct-1992 #text_change 08-Oct-1999

C:Accession: A36811

R:Albrecht, J

submitted to the EMBL Data Library, January 1992

A:Description: Primary structure of the herpesvirus saimiri genome.

A:Reference number: A36806

A:Accession: A36811

A:Molecule type: DNA

A:Residues: 1-797 <ALB>
 A:Cross-references: GB:X64346; NID:960320; PIDN:CAA45671.1; PID:960369
 R:Albrecht, J.C.; Nicholas, J.; Biller, D.; Cameron, K.R.; Blesinger, B.; Newman, C.;
 J. Virol. 66, 5047-5058, 1992
 A:Title: Primary structure of the herpesvirus saimiri genome.
 A:Reference number: A37309; MUID:9233688; PMID:1321287
 A:Contents: annotation; protein-coding frames
 A:Note: neither protein nor nucleotide sequence is given
 C:Genetics:
 A:Gene: 48

Alignment Scores:

Pred. No.:	Length:	Score:	Matches:	Conservative:	Best Local Similarity:	Mismatches:	Indels:	Query Match:
5.17	797	78.00	31	12	44.79%	33	20	11.57%
2	5	11.57%	20	20				

US-09-936-737a-1 (1-375) x A36811 (1-797)

QY 94 GCGAAGAAATATACAGACTTGCATAATCTTTAAGAATGCTCTGATCTGACGA 153
 Db 382 AlaAsnGluLys-----GluTyrLysLysIleIleAspLysSerAspAspArgAsp 399
 QY 154 TGCAAAAAACAATGTTTCAAGACGAGTACTGATCGTTTGAAGACACGCTCAC 213
 Db 400 ArgAspLysAspGluTyrGluLeuGlu-----Asn 409
 QY 214 AAGAAATGTTACTACAAAT-----GTGTTATGCTGAAGATTGACCAAGAA 261
 Db 410 GluGlu-----TyrAsnArgAspGluGluGluAspGluGluAspGluGluAspGlu 427
 QY 262 AAATTTGTTGTCGACGAAACTTCACGAAATATATTGACACACGCGAGGTTAAGAT 321
 Db 428 LysAspGluLysGluGluGluGluGluAspGluGluAspGluGluGluGluAsp 447
 QY 322 GCAGTATATGCGCAGTACAGTGCAGCTCATGAAAGTTGATGAA 369
 Db 448 GluGluGlu-----AspGluGluAspGluGluGluGluGluGluGluGlu 461

RESULT 14

T15129

hypothetical protein ZC328.5 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 04-Mar-2000

C:Accession: T15129

R:Mansley, P.

submitted to the EMBL Data Library, April 1997

A:Description: The sequence of C. elegans cosmid ZC328.

A:Reference number: Z18298

A:Accession: T15129

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-191 <WAM>

A:Cross-references: EMBL:AF000194; NID:G1946990; PID:G1946994; PIDN:AAB52895.1; GSPDB

A:Experimental source: strain Bristol N2; clone ZC328

C:Genetics:

A:Gene: CESP:ZC328.5

A:Map position: 1

A:Introns: 20/3; 124/3

C:Superfamily: Caenorhabditis elegans hypothetical protein ZC328.5

Alignment Scores:

Pred. No.:	Length:	Score:	Matches:	Conservative:	Best Local Similarity:	Mismatches:	Indels:	Query Match:
5.4	191	77.50	23	15	35.51%	28	41	11.50%
2	4	11.50%	41	41				

US-09-936-737a-1 (1-375) x T15129 (1-191)


```

QY 1 ATGAATATTTCTGATTCCTTCCTTGCCTGCAAGCTTGCTGATC----- 48
    |||:::  ::  |||  |||  ::  |||
Db 1 MetArgheglnIlePheThrPheLeuAlaLeuAlaAlaIleasnIleGlylu 20
QY 48 ----- 48
Db 21 GlyAlaIysProCysAlaIysArgProThrThrGluProProSerThrThrVal 40
QY 49 ---TCAACTACTCTTCAGAGAAAGTGAAGATTGTTGACGTTTACGCGAACAAGAA 105
    |||:::  ::  |||  |||  ::  |||
Db 41 GluSerSerThrThrArgSpsSerLysAlaProCysMetAlaValPheGluAsnMetLys 60
QY 106 -----TATACAGACTTCGATTAATCTTTAAGACGCCCTGATCTGAC 150
    ::  |||:::  ::  |||:::  ::  |||
Db 61 GluGluGluAspGlyPheGluAspTyrAlaMetGlyValThrThrAlaIleServalPro 80
QY 151 GAATGCAAAAACATG-----TTCAGACGCGAG 180
    |||  |||  |||  |||  |||  |||
Db 81 GluCysSerLysThrCysGluThrAspValLysCysAlaTyrSerPhePheLysThrAsn 100
QY 181 TACTGCTACATCGTTTGA 201
    |||  |||  |||  |||  |||
Db 101 LysCysTyr--AlaPheGlu 106

```

RESULT 15

G90594

hypothetical protein MYPU_6630 [imported] - Mycoplasma pulmonis (strain UAB CTIP)

C:Species: Mycoplasma pulmonis

C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001

C:Accession: G90594

R:Chambaud, I.; Hellig, R.; Ferris, S.; Barbe, V.; Samson, D.; Gallison, F.; Moszer, I.;

Nucleic Acids Res. 29, 2145-2153, 2001

A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm

A:Reference number: A99512; MUID:21267165; PMID:11353084

A:Accession: G90594

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-555 <KUR>

A:Cross-references: GB:AL445566; PID:g14090078; PIDN:CAC13836.1; GSPDB:GN00153

A:Experimental source: strain UAB CTIP

C:Genetics:

A:Gene: MYPU_6630

A:Genetic code: SGC3

Alignment Scores:

```

Pred. No.: 6.47 Length: 555
Score: 77.00 Matches: 37
Percent Similarity: 38.89% Conservative: 19
Best Local Similarity: 25.69% Mismatches: 62
Query Match: 11.42% Indels: 26
DB: 2 Gaps: 4

```

US-09-936-737a-1 (1-375) x G90594 (1-555)

```

QY 13 TTGATTTCTTCCTTGCCTGCAAGCTTGATCTCAACTTCTTCAGAGACGT 72
    |||:::  |||  |||  |||  |||  |||
Db 53 LeuValAspPheLeuThrLysLysAsnIleAsnIleSerThrIleLysIleAspAspGlu 72
QY 73 GAAGATTGTTGACGTTTACGCGAACAATAATACAGACTTCGATTAATCTTTAAG 132
    |||  |||  |||  |||  |||  |||
Db 73 GluIleSerLeuSerAspTyrThrLysValLysGlnGlyAspPheAspLysLysAlaAsp 92
QY 133 AAG-----TCTCTGATCTTGACGAATGCAAA-----AAA 162
    |||  |||  |||  |||  |||  |||
Db 93 LysLeuAsnPheIleSerArgLysAsnAspGluIleLysThrLysLeuAspAsnAsnLys 112
QY 163 ACATGTTCAAGAGC---GAGTACTGCTTACATCGTTTGAAGACACGTCACACAGGAA 219
    |||  |||  |||  |||  |||  |||
Db 113 ThrIleLysGlnHisGlnSerLeuAspGlnPheValAspGluIleIleGluLys 132
QY 220 TGTTACTACATGTCGTTGAT----- 240
    :::  |||  |||  |||  |||  |||
Db 133 AsnSerLeuSerLeuValAspLysLeuLysGluPheLysGluLysIleAspTyrPheAsn 152

```

```

QY 241 -----GCTGAGAGTTAGACCAAGAAATTTGTTGTCGACGAAACTTCACGGAAT 294
    :::  |||  |||  |||  |||  |||
Db 153 GluLysSerGlnGluIleAsnTyrPargSerPheTrpGluValLysAsnProAlaAspAsn 172
QY 295 TATTGACAGACTGCGAGGTTAAGATGCGAGTAAATGCGCGACGTTACAGTACAGTCA 354
    |||  |||  |||  |||  |||  |||
Db 173 IleArgPheAspProGluLysIleAspTyrAspAsnAlaLysGluIleGlyLysLeuLys 192
QY 355 GATGAAGTTGAT 366
    :::  |||  |||  |||  |||  |||
Db 193 AsnLysIleAsp 196

```

Search completed: February 25, 2003, 02:29:12
 Job time : 27 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic acid - protein search, using frame_plus_n2p model

Run on: February 25, 2003, 01:50:35 ; Search time 12.5 Seconds

(Without alignments)
2488.580 Million cell updates/sec

Title: US-09-936-737A-1

Perfect score: 6/4

Sequence: 1 atgaagatcttcgtatc.....atgaagatcttcgaatataa 375

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Zgapop 6.0 , Zgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+np2 model -DEV=xlh
-O=/cgm2.1/USP/seq/US09936737/runt_14022003_100557_1389/app_query.fasta.1.519
-DB=SwissProt_40 -QPM=fastan -SUFFIX=ssp -MINMATCH=0.1 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=human40.cdt -LIST=45
-DOCALIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09936737@cgm2.1.5.6 @runat_14022003_100557_1389 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-NARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -YGAPOP=6 -YGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	95.5	14.2	147	1	LAPP_HAEOP
2	82.5	12.2	412	1	FKB1_SPOFR
3	79.5	12.2	507	1	YGA6_YEAST
4	79.5	11.8	1038	1	BMR2_MOUSE
5	78	11.6	797	1	VG48_HSVSA
6	76	11.3	559	1	TCPG_JETPY
7	75.5	11.2	419	1	ICBC_MOUSE
8	75.5	11.2	491	1	P2C1_CAEEL
9	75	11.1	360	1	YORA_CAEEL
10	75	11.1	2410	1	MOK1_SCHPO
11	74.5	11.1	146	1	PA21_CANPA
12	74.5	11.1	146	1	PA21_PIG
13	74.5	11.1	451	1	Z222_HUMAN
14	74	11.0	296	1	YJ23_AOUAE
15	74	11.4	600	1	SP96_DICDI
16	73.5	11.0	4465	1	DYHC_ANTCR
17	73.5	10.9	208	1	RK1_PEA
18	73.5	10.9	562	1	CHIT_YEAST

C	19	73.5	11.3	636	1	YNR6_YEAST	P53882	saccharomyc
C	20	73.5	10.9	2339	1	RPCL_PLAFA	P27625	plasmodium
C	21	73	11.2	551	1	YGF1_YEAST	P53214	saccharomyc
C	22	73	10.8	1405	1	DPOA_SCHPO	P28040	schizosacch
C	23	72.5	11.1	1802	1	HKR1_YEAST	P41809	schizosacch
C	24	72	11.1	338	1	YD49_THENA	P41809	schizosacch
C	25	72	10.7	587	1	YKRO_YEAST	P09170	thermotoga
C	26	71.5	10.6	146	1	PA21_RAT	P34248	saccharomyc
C	27	71.5	10.6	156	1	CB20_HUMAN	P04055	saccharomyc
C	28	71.5	10.6	1177	1	Y307_MYCE	P52288	rattus norv
C	29	71	10.9	110	1	YCT1_YEAST	P47549	mycoplasma
C	30	71	10.9	373	1	YVW2_YEAST	P37265	saccharomyc
C	31	71	10.5	782	1	YVW2_YEAST	P34735	pichia angu
C	32	71	10.5	2555	1	FAFY_HUMAN	P40157	saccharomyc
C	33	70.5	10.5	78	1	MOH1_CANGG	000507	h probabile
C	34	70.5	10.5	346	1	YFEL_YEAST	P81034	cancer pagu
C	35	70.5	10.5	748	1	Y875_MENJA	P41834	saccharomyc
C	36	70.5	10.5	793	1	MDOB_DROME	Q58285	methanococ
C	37	70	10.4	763	1	MDOB_DROME	Q24568	drosophila
C	38	70	10.4	1058	1	S185_YEAST	P39401	escherichia
C	39	69.5	10.7	264	1	UL31_HSV6U	P40856	saccharomyc
C	40	69.5	10.3	733	1	YU08_YEAST	P28865	human herpe
C	41	69.5	10.7	1468	1	N153_RAT	P45995	saccharomyc
C	42	69	10.2	644	1	KNG_HUMAN	P45991	rattus norv
C	43	69	10.2	1056	1	RIQ1_YEAST	P01042	homo sapien
C	44	68.5	10.2	591	1	CAIX_RAT	P40395	saccharomyc
C	45	68.5	10.2	610	1	CAIX_HUMAN	P35565	rattus norv
							O14967	homo sapien

ALIGNMENTS

RESULT 1
LAPP_HAEOP
ID LAPP_HAEOP STANDARD: PRT: 147 AA.
AC Q01747;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE Anti-platelet protein precursor.
GN LAPP.
OS Haementeria officialis (Mexican leech).
OC Eukaryota; Metazoa; Annelida; Clitellata; Hirudinea; Hirudinea;
OC Rhynchobdellida; Glossiphoniidae; Haementeria.
OX NCBI_TaxID=6410;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 60-91 AND 123-139.
RC TISSUE=Salivary gland;
RX MEDLINE=92202247; PubMed=155198;
RA Keller P.M., Schultz L.D., Condra C., Karczewski J., Connolly T.M.;
RT "An inhibitor of collagen-stimulated platelet activation from the
RT salivary glands of the Haementeria officialis leech. II. Cloning of
RT the cDNA and expression."
RT J. Biol. Chem. 267:6899-6904(1992).
CC - FUNCTION: AN INHIBITOR OF COLLAGEN-STIMULATED PLATELET
CC AGGREGATION, DENSE GRANULE RELEASE AND SEROTONIN RELEASE.
CC - SUBCELLULAR LOCATION: Secreted.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
DR EMBL, M81489; AAA29194.1; -
KW Signal.
FT SIGNAL
FT CHAIN
SQ SEQUENCE 147 AA; 15908 MW; 75A5511374AA42E CRC64;
Alignment Scores: 0.0379 Length: 147
Pred. No.: 147

Score: 95.50 Matches: 31
Percent Similarity: 36.07% Conservative: 13
Best Local Similarity: 25.41% Mismatches: 31
Query Match: 14.17% Indels: 47
DB: 1 Gaps: 4

US-09-936-737a-1 (1-375) x LAMP_HADEF (1-147)

QY 1 ATGAAGATTTCTGATTTCTCTCTGCGAAGCTTGCATCTCA----- 51
DB 1 Metasnerheleupheserleualacys-----Serleulevalalaleproala 18
QY 51 ----- 51
DB 19 lleseralaglnaspgluasapalaglyalaglyaspcluthrsercluglglunasp 38
QY 52 ACTACTCTCTCAGAGAA----- 69
DB 39 Thrthnglyseraspcluthrproserthrglygllyaspclgluasnglunlu 58
QY 70 -----CCTGAAGATTTGTCAGCTTTTACGCAAGCAAGAAATATACAGAC 114
DB 59 ThrllthrAlaglyaspnglunaspCysTyrIleleuInlleasnThrgluthrnsnglucystyr 78
QY 115 TTCGATTAATCTTTTAAAGAGTCTCTGATCTGACGATGCAAAAAACATGTTCAAG 174
DB 79 AsnleuleuthrlysthrthgluphethrSerValaspclucysargylsmtcysglunlu 98
QY 175 ACGAGATAC-----TGTACATCTGTTTTCAGACACGCGTCAAGCAAGATGTTAC 225
DB 99 SerAlavalagluProserCysTyrIleleuInlleasnThrgluthrnsnglucystyr 118
QY 226 TACAAAT 231
DB 119 Argasn 120

RESULT 2
FKB4_SPOFR STANDARD: PRT: 412 AA.

ID FKB4_SPOFR
AC 026486;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE 46 kDa FK506-binding nuclear protein (Peptidyl-prolyl cis-trans isomerase) (PPIase) (EC 5.2.1.8).
OS Spodoptera frugiperda (Fall armyworm).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;
OC Diptera; Noctuidae; Noctuidae; Amphipyrinae; Spodoptera.
NCBI_TaxID=7108;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95074110; PubMed=7527037;
RA Alnemri E.S., Fernandes-Alnemri T., Pomeroy K., Robertson N.M., Dudley K., Dubois G.C., Litwack G.;
RT "FKBP46, a novel sf9 insect cell nuclear immunophilin that forms a protein-kinase complex."
RL J. Biol. Chem. 269:30828-30834(1994).
CC -1- FUNCTION: BINDS TO, AND IS INHIBITED BY FK506 AND RAPAMYCIN. BINDS DOUBLE-STRANDED DNA IN VITRO. PPIASES ACCELERATE PROTEIN FOLDING.
CC -1- CATALYTIC ACTIVITY: CIS-TRANS ISOMERIZATION OF PROLINE IMIDIC PEPTIDE BONDS IN OLIGOPEPTIDES.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- PTM: PHOSPHORYLATED BY A NUCLEAR KINASE IN THE PRESENCE OF MG2+ AND ATP.
CC -1- SIMILARITY: BELONGS TO THE FKBP-TYPE PPIASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation in the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

or send an email to license@sib-sib.ch.

CC -----
CC EMBL: U15038; AA58962.1; -
DR HSSP: P20071; 1FKJ.
DR InterPro: IPR001179; FKB_PPIase.
DR Pfam: PF00254; FKB_PPIase_1.
DR PROSITE: PS00453; FKB_PPIase_1; FALSE_NEG.
DR PROSITE: PS00454; FKB_PPIase_2; 1.
DR PROSITE: PS00509; FKB_PPIase_3; 1.
KW Isomerase; Rotamase; Nuclear protein; DNA-binding; Phosphorylation.
FT DOMAIN 91 112 Asn/Glu-Rich (HIGHLY ACIDIC).
FT DOMAIN 120 145 Lys-Rich (BASIC).
FT DOMAIN 152 216 Asp/Glu-Rich (HIGHLY ACIDIC).
FT DOMAIN 219 302 Lys-Rich (BASIC).
FT DOMAIN 324 412 PPIase, FKB-TYPE.
SQ SEQUENCE 412 AA; 45610 MW; F2A69159AER4FE22 CRC64;

Alignment Scores:
Pred. No.: 0.896 Length: 412
Score: 82.50 Matches: 32
Percent Similarity: 41.74% Conservative: 16
Best Local Similarity: 27.83% Mismatches: 40
Query Match: 12.24% Indels: 27
DB: 1 Gaps: 5

US-09-936-737a-1 (1-375) x FKB4_SPOFR (1-412)

QY 43 CTGATCTCAACTACTCTCTCAGAGACGTCGATGTTGCGAGCTTTACGCAACAGA 102
DB 116 LeuValProAlaLysAsnLysArgLysLeuLysAlaAsnAlaLarAlaAsnLys 135
QY 103 AAATATACAGACTCTGATTAATCTTTTAAAGATCC-----TCGATCTT 147
DB 136 LysAlaLysProAspLysLysAlaGlyLysAsnSerAlaProAlaAlaGluSerAspSer 155
QY 148 GACGATGCAAAAAAACATGTTTCAAGACGAGTCTGATCTGATCTTTTGAAGACAGC 207
DB 156 AspAspAspAsp-----GluAspGln 162
QY 208 GTCACACAGGATGTTACTACATGTCGTTGATGTCGAGATTTAGCCAAAGAAATTT 267
DB 163 LeuLys-----PheLeuAspGlyGluAspLysPheAspThrAspGlu 176
QY 268 GTTGTCACAGAAACTCTCAGCAAGAAATTTTTCAGACAGCTGCGATGAAATGACAGT 327
DB 177 --AsnAspGluSerPheLysMetAsn-----ThrSerAlaGluGlyAspAspSerAsp 193
QY 328 AATGCGCAGCTGACGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 372
DB 194 GluGluAspAspAspAspGluAspGluGluAspGluGluAspAspAsp 208

RESULT 3
YG46_YEAST STANDARD: PRT: 507 AA.

ID YG46_YEAST
AC P53301;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 52.8 kDa protein in BUB1-HIP1 intergenic region.
GN YGR189C OR G7553.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=97279231; PubMed=9133739;
RA Arroyo J., Garcia-Gonzalez M., Garcia-Saez M.I., Sanchez-Perez M., Nombela C.;
RT "DNA sequence analysis of a 23,002 bp DNA fragment of the right arm of Saccharomyces cerevisiae chromosome VII.";
RL Yeast 13:357-363(1997).

CC	Eukaryota, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
CC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxId=10090;		
RN	SEQUENCE FROM N.A.		
RP	MEDLINE-97350808; PubMed=9207184;		
RX	heppu H., Minowa O., Miyazono K., Kawabata M.;		
RT	"cDNA cloning and genomic organization of the mouse BMP type II		
RT	receptor.";		
RL	Biochem. Biophys. Res. Commun. 235:499-504(1997).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RA	Wittek G.B., Koenig B.B., Ting J., Tiesman J.P., Limberg A.L.,		
RA	Grant R.A., Begley K.B., Rosenbaum J.S.;		
RT	"Identification of BMP receptor complexes with differential signaling		
RT	properties and ligand binding profiles.";		
RL	Submitted (Dec-1997) to the EMBL/GenBank/DBJ databases.		
CC	-1- FUNCTION: BINDS TO BMP-7, BMP-2 AND, LESS EFFICIENTLY, BMP-4.		
CC	-1- BINDING IS WEAK BUT ENHANCED BY THE PRESENCE OF TYPE I RECEPTORS		
CC	FOR BMPs.		
CC	-1- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.		
CC	-1- SUBUNIT: HETERODIMERIZE WITH TYPE-I RECEPTOR.		
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein.		
CC	-1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.		
CC	TGF β RECEPTOR SUBFAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its use		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (see http://www.1sb-sib.ch/announce/		
CC	or send an email to license@sib-sib.ch).		
CC	-----		
DR	EMBL; AF003942; AAB63042.1; -		
DR	EMBL; U78048; AAB87638.1; -		
DR	MGD; MGI:1095407; Bmpr2.		
DR	InterPro; IPR0004472; Activin_rec.		
DR	InterPro; IPR000719; Euk_pkinase.		
DR	InterPro; IPR002290; Ser_thr_pkinase.		
DR	Pfam; PF00069; pkinase; 1.		
DR	Pfam; PF01064; Activin_recpt; 1.		
DR	Prodom; PD000001; Euk_pkinase; 1.		
DR	ProSite; PS00107; PROTEIN KINASE ATP; 1.		
DR	ProSite; PS00108; PROTEIN KINASE ST; FALSE-NEG.		
DR	ProSite; PS50011; PROTEIN_KINASE_DOM; 1.		
RW	Receptor; Transfease; Serine/threonine-protein kinase; ATP-binding;		
RW	Transmembrane; Glycoprotein; Signal.		
FT	SIGNAL		
FT	1		
FT	27		
FT	1038		
FT	CHAIN		
FT	27		
FT	1038		
FT	DOMAIN		
FT	27		
FT	150		
FT	171		
FT	TRANSSEM		
FT	151		
FT	171		
FT	DOMAIN		
FT	172		
FT	1038		
FT	DOMAIN		
FT	203		
FT	504		
FT	NP_BIND		
FT	209		
FT	217		
FT	BINDING		
FT	230		
FT	230		
FT	333		
FT	ACT_SITE		
FT	191		
FT	194		
FT	DOMAIN		
FT	547		
FT	550		
FT	DOMAIN		
FT	610		
FT	618		
FT	DOMAIN		
FT	901		
FT	908		
FT	CARBOHYD		
FT	55		
FT	55		
FT	CARBOHYD		
FT	110		
FT	110		
FT	CARBOHYD		
FT	126		
FT	126		
SO	SEQUENCE		
SO	1038 AA; 115019 MW; 4106945DC63250E1 CRC64;		
Alignment Scores:			
Pred. No.:	1.94	Length:	1038
Score:	79.50	Matches:	27
Percent Similarity:	40.46%	Conservative:	26
Best Local Similarity:	20.61%	Mismatches:	35
Query Match:	11.80%	Indels:	43

Alignment Scores:

Pred. No.: 4.3
 Score: 76.00
 Percent Similarity: 42.34%
 Best Local Similarity: 26.13%
 Query Match: 11.28%
 DB: 11.28%
 Length: 559
 Matches: 29
 Conserved: 18
 Mismatches: 30
 Indels: 35
 Gaps: 4

US-09-936-737a-1 (1-375) x TCGP_TEMP (1-559)

QY 56 CTCTTCAGAGAGAGAGATGCTGAGCTTTTACGCGACAGAGAAATATACAGACT 115
 DB 307 LeuLeuLysLysLysLysValSerValLleArgValValArgLysThrAspAsn 323
 QY 116 TCGATTAATCTTTAAGAAAGCTCTGTGATCTTGACAGATGCAAAAAACATGTTCAAGA 175
 DB 324 -----
 QY 176 CGAGTACTGCTACATC-----GTTTGAAGACACGCTACAGAG 216
 DB 329 ValSerGlyAlaThrLleValAsnArgProGluGluLleGlnLysSerAspValGlyLys 348
 QY 217 GAATGTTACTACAAATGCTGTTGATGAGAGAGTGAACCAAGAAATTTGTCGAC 276
 DB 349 LysCys-----GlyLeuPheGluValLysLeuLleGlyAsp 360
 QY 277 GAAACCTTCACGAAATTTATTTGACAGACTGCGAGGTAAAGATGAGTATGCGCA 336
 DB 361 GluTyrPheThr-----PheMetThrGluCysGluAsnProGluValLysCysLleLle 378
 QY 337 GGTACAGGTACAGAGTACAGATGAGTGAATGAGAA 369
 DB 379 LeuArgGlyAlaSerLysAspValLleuAsnGlu 389

RESULT 7

ICEC_MOUSE
 ID ICEC_MOUSE STANDARD: PRT; 419 AA.

AC 008736;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Caspase-12 precursor (EC 3.4.22.-).
 GN CASP12.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C3H/An.
 RX MEDLINE=97190206; PubMed=9038361;
 RA van de Craen M., Vandenaabeele P., Declercq W., van den Brande I.,
 RA Fiers W., Mollemans F., Schotte P., van Criekinge W., Beyaert R.,
 RA Fiers W.;
 RT "Characterization of seven murine caspase family members.";
 RL FEBS Lett. 403:61-69(1997).
 CC -1- FUNCTION: INVOLVED IN THE ACTIVATION CASCADE OF CASPASES
 CC RESPONSIBLE FOR APOPTOSIS EXECUTION (BY SIMILARITY).
 CC -1- SUBUNIT: HETERODIMER OF TWO SUBUNITS (POTENTIAL).
 CC -1- TISSUE SPECIFICITY: MAINLY EXPRESSED IN SKELETAL MUSCLE AND LUNG.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C14.
 CC -1- SIMILARITY: CONTAINS 1 CARD DOMAIN.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as the content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 DR EMBL: Y13090; CAA73532.1; -
 DR HSP: P29466; IICE.

DR MEROPS: C14.013; -
 DR MGD: MGI:1312922; Casp12.
 DR InterPro: IPR001315; CARD.
 DR InterPro: IPR002398; ICE.
 DR InterPro: IPR002138; ICE_p10.
 DR InterPro: IPR001309; ICE_p20.
 DR Pfam: PF00655; ICE_p10; 1.
 DR Pfam: PF00656; ICE_p20; 1.
 DR PRINTS: PR00376; ILIBENZYM.
 DR SMART: SM00114; CARD; 1.
 DR SMART: SM00115; CAS; 1.
 DR PROSITE: PS50209; CARD; 1.
 DR PROSITE: PS01122; CASPASE_CYS; 1.
 DR PROSITE: PS01121; CASPASE_HIS; 1.
 DR PROSITE: PS02027; CASPASE_P10; 1.
 DR PROSITE: PS02028; CASPASE_P20; 1.
 DR Hydrolyse: Thiol protease; Apoptosis; zymogen.
 FT PROPEP 1 ? ?
 FT CHAIN 1 ? ?
 FT PROPEP 2 ? ?
 FT CHAIN 2 ? ?
 FT DOMAIN 1 92
 FT ACT_SITE 250 250
 FT ACT_SITE 298 298
 SQ SEQUENCE 419 AA; 47853 MW; B94B0FED16B1CB40 CRC64;

Alignment Scores:

Pred. No.: 4.75
 Score: 75.50
 Percent Similarity: 54.79%
 Best Local Similarity: 35.62%
 Query Match: 11.20%
 DB: 11.20%
 Length: 419
 Matches: 26
 Conserved: 14
 Mismatches: 14
 Indels: 19
 Gaps: 6

US-09-936-737a-1 (1-375) x ICEC_MOUSE (1-419)

QY 193 GTTGTGAAGACAGCTCAACAAGAAATGTTACTACATGCTGTTGATGAGTGAAGAGTTA 252
 DB 26 ValPheAspSerLeuValGluLys-----AsnValLleuAsnGlyAspLleu 41
 QY 253 -----GACCAAGAAATTTGTTGTCGAC-----GAAACCTTCACGAAAT 294
 DB 42 LeuLysLleGlyLysSerLysPheLleuAsnLysAlaGluAsnLysValGluAsn 61
 QY 295 TATTTG-----ACAGCTGCGAGGTAA--GATGCAGGTAATGCGCAGTACAGT 345
 DB 62 PheLeuGluLysThrAspMetAlaGlyLysLlePheAlaGlyHisLleAlaAsnSerGln 81
 QY 346 GAGCAG-----TCAGATGAGTTGAT 366
 DB 82 GluGlnLeuSerLleuGlnPheSerAsnAspLysAsp 94

RESULT 8

P2C1_CAEEL
 ID P2C1_CAEEL STANDARD: PRT; 491 AA.

AC P49595;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Probable protein phosphatase 2C F42G9.1 (EC 3.1.3.16) (PP2C).
 GN F42G9.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peleoderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Tatch A., Waterston R.;
 RA Submitted (MAR-1996) to the EMBL/Genbank/DBJ databases.
 RL -1- CATALYTIC ACTIVITY: A phosphoprotein + H(2)O -> a protein +
 CC phosphate.
 CC -1- COFACTOR: BINDS 2 MAGNESIUM OR MANGANESE IONS (BY SIMILARITY).
 CC

RP SEQUENCE FROM N.A.
RC STRAIN-Bristol N2;

RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
PA Wood V., Gilliam R., Rajadream M.A., Lyne M., Lyne R., Stewart A.

RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
PA Wood V., Gilliam R., Rajadream M.A., Lyne M., Lyne R., Stewart A.

OY 259 GAAAAATT-----CITTCGAGCAAACTTCCAGGAAT-----TAFTTGACA 303
||| ||| :::::||:|||| |
Db 706 TyAlAphelyLeuTyValProglUGluAspheIleGLuasnasPproIlelthr 725

OY 304 GA CTGGACAGGTFAAGATGCAGGTACTGCCGCGCATCACAGTAGCAGCTAGCAT 357
:::
Db 726 SerLeuthrPROGUHISaspALargyvalValalasecDlyasnGUlleASP 743

RESULT 11
ID PA21_CANFA STANDARD; PRT: 146 AA.
PA21_CANFA

AC P06596;
DT 01-JAN-1988 (Rel. 06, Created)
DD 01-JAN-1998 (Rel. 06, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DN Phospholipase A2 precursor (EC 3.1.1.4) (Phosphatidylcholine 2-acylhydrolase) (Group IB phospholipase A2).
GN PLAZSIB.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eutelestomi;
OX Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.

CX NCBI_TaxID=9615;

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86233862; PubMed=3754861;
RA O'hara O., Tamaki M., Nakamura E., Tsuruta Y., Fujii Y., Shin M.,
RZ "Dog and rat pancreatic phospholipases A2: complete amino acid sequences deduced from complementary DNAs."
RN J. Biochem. 99:733-739(1986).

[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90306027; PubMed=2142076;
RA Kerfelec B., Laforge K.S., Vasilakides P., Pulgoverver A., Scheele G.A.;
RZ "Isolation and sequence of the canine pancreatic phospholipase A2 gene." ;
RN Eur. J. Biochem. 190:299-304(1990).

[3]
RP SEQUENCE FROM N.A.
RX MEDLINE=87175472; PubMed=3562437;
RA Kerfelec B., Laforgue K.S., Pulgoverver A., Scheele G.A.;
RZ "Primary structures of canine pancreatic lipase and phospholipase A2 messenger RNAs." ;
RN Pancreas 1:430-433(1986).

RL -1- FUNCTION: PA2 catalyzes the calcium-dependent hydrolysis of the
2-acyl groups in 3-sn-phosphoglycerides.
CC -1- CATALYTIC ACTIVITY: Phosphatidylcholine + H(2O) = 1-
acylglycerophosphocholine + a fatty acid anon.
CC -1- COFACTOR: Binds 1 calcium ion per subunit (by similarity).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE A2 FAMILY.

CC This SWISS-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at its European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or send an email to license@lsb-sib.ch).

-- -- --
EMBL: D00035; BAU0023.1; .
DR EMBL: M35301; AAA30883.1; .
DR PIR: A24392; PSDG.
DR PIR: S11316; S11316.
DR HSSP: P00592; AP2P.
DR InterPro: IPR001211; PhospholipaseA2.
DR Pfam: PF00068; Phoslipd_1.
DR PRINTS: PR00389; PRRHIPASER2.
DR ProDom: PD000303; PhospholipasaeA2; 1.
DR SMART: SM00085; PA2c; 1.
DR PROSITE: PS00118; PA2_HIS_1.
DR PROSITE: PS00119; PA2_ASP_1.
KW hydrolase, lipid degradation, Calcium, Pancreas, Signal.

Score:	5.63	Length:	146
Percent Similarity:	74.50	Matches:	29
Best Local Similarity:	38.33%	Conservative:	17
Query Match:	24.17%	Mismatches:	39
DB:	11.05%	Indels:	35
	1	Gaps:	5

Db
1 MetIvSphelenvat
::: ||||| |||||::: |||||
atgacccttcttggccctgcgaagcttgctgatctcaactacttcttcagagaa----- 6

70 -----CGTGAAGATTGTTGGACGCTGGT-----ThrValAlaAlaIagIugIygly I

19 IleserProargalavaiTirPGlnPhearTAsenwctrr-----11

112 GACTTCGATAAATCTTTAG-----gasmucellicyscsthrlleprogluser 38

39 AspProLeuLysAspTyrAsnAspTyrGlyCysTyrCysGlyLeuGlyVal¹⁴-----AAGTCCTCT 14

142 GATCTTGACGAAATGCMAAAAAACATGTTTCAAGACGAGTACTGCTACATCGTTTTGA 20
:::||||| ||| |||
50 - - - - - ||| |||

[illegible]

75 ----- : : : : :
ATGCGGCACACAGGAATGTACTACAATGTCGTTGATGCTGAAGAGTTAGACCAGA

262 AAATTTGTTCTCGACGAAGAACTTTCACC...
-----SerGIuaIalLyslSylLeuAspSerCys 83

84 LysPheLeuLeuAspAsnProTyrThrIvsI]etvscm... 321

21-PTG 103

PA21_PIG
P00592;
STANDARD;
PRT; 146 AA.

21-JUL-1986	(Rel. 01, Created)
01-JAN-1988	(Rel. 06, Last sequence updated)
15-JUN-2003	

Phospholipase A2, major isoenzyme precursor (EC 3.1.1.4)
(Phosphatidylcholine 2-acylhydrolase)
Phospholipase A2, last annotation update
Phospholipase A2, major isoenzyme precursor (EC 3.1.1.4)
(Phosphatidylcholine 2-acylhydrolase)
Phospholipase A2, last annotation update

PLA2G1B. (EC 3.1.1.4)
Sus scrofa (Pig).
Group IB phospholipase A2.

Eukaryota; Metazoa; Chordata; Cranata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Sauria

SEQUENCE FROM N 2
[1]
nccl_taxid=9823;
Sulda; Sus.

TISSUE=Pancreas;
 MEDLINE=87231066; PubMed=33067700

de Geus P., van den Bergh C.J., Kuipers O., Verheij H.M.,
Hoekstra W.P.M., de Haas G.H.

Expression of porcine pancreatic phospholipase A₂. Generation of

RT RT active enzyme by sequence-specific cleavage of a hybrid protein from
RT Escherichia coli." ;
RL Nucleic Acids Res. 15:3743-3759(1987).
RN [2]
RP SEQUENCE FROM N.A.
RP TISSUE-Pancreas;
RX MEDLINE-87132925; PubMed-3028739.
RA Sellmeyer J.J., Randall T.L., Yamataka M., Johnson L.K.;
RT "Pancreatic phospholipase A2: Isolation of the human gene and
RT from porcine pancreas and human lung." ;
RN DNA 5:519-527(1986).
RN [3]
RP SEQUENCE OF 16-146.
RX MEDLINE-71014235; PubMed-5528841;
RA de Haas G.H., Slotboom A.J., Bensen P.P.M., van Deenen L.L.M.,
RT "Studies on phospholipase A and its zymogen from porcine pancreas. I.
RT The complete amino acid sequence." ;
RL Biochim. Biophys. Acta 221:31-53(1970).
RN [4]
RP REVISIONS.
RX MEDLINE-7722092; PubMed-884127;
RA Puljk W.C., Verheij H.M., de Haas G.H.;
RT "The primary structure of phospholipase A2 from porcine pancreas. A
RT bioinvestigation." ;
RL Biochim. Biophys. Acta 492:254-259(1977).
RN [5]
RP DISULFIDE BONDS.
RX MEDLINE-71014236; PubMed-4919729;
RA de Haas G.H., Slotboom A.J., Bensen P.P.M., Nieuwenhuizen W.,
RT "Studies on phospholipase A and its zymogen from porcine pancreas.
RT II The assignment of the position of the six disulfide bridges." ;
RL Biochim. Biophys. Acta 221:54-61(1970).
RN [6]
RP ACTIVATION OF LYS-78.
RC TISSUE-Pancreas;
RX MEDLINE-89255488; PubMed-2498336;
RA Tomaselli A.G., Hui J., Fisher J., Zuercher-Neely H., Reardon H.M.,
RT "Oxidation and activation of porcine pancreatic phospholipase A2
RT via substrate level acylation of lysine 56." ;
RL J. Biol. Chem. 264:10041-10047(1989).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS), AND ACTIVE AND BINDING SITES.
RX MEDLINE-83268704; PubMed-6876174;
RA Dijkstra B.W., Renssester R., Kalk K.H., Hol W.G.J., Drenth J.;
RT "Structure of porcine pancreatic phospholipase A2 at 2.6-A resolution
RT and comparison with bovine phospholipase A2." ;
RL J. Mol. Biol. 168:163-179(1983).
RN [8]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).
RX MEDLINE-91073405; PubMed-2254938;
RA Thunnissen M.M.G.M., Kalk K.H., Drenth J., Dijkstra B.W.;
RT "Structure of an engineered porcine phospholipase A2 with enhanced
RT activity at 2.1-A resolution. Comparison with the wild-type porcine
RT J. Mol. Biol. 216:425-439(1990).
RN [9]
RP STRUCTURE BY NMR.
RX MEDLINE-91175768; PubMed-2007145;
RA Dekker N., Peters A.R., Slotboom A.J., Boelens R., Kaptein R.,
RT de Haas G.H.;
RT "Porcine pancreatic phospholipase A2: sequence-specific 1H and 15N
RL NMR assignments and secondary structure." ;
RN Biochemistry 30:3135-3147(1991).
RN [10]
RP STRUCTURE BY NMR.
RX MEDLINE-95393214; PubMed-7664098;
RA van den Berg B., Tessari M., Boelens R., Dijkman R., de Haas G.H.,
RT "NMR structures of phospholipase A2 reveal conformational changes
RT during interfacial activation." ;

FT	CA_BIND	54	54	VIA CARBOXYL OXYGEN.
FT	CA_BIND	71	71	
FT	LIPID	78	78	N-OCTANOATE.
FT	HELIX	24	34	
FT	TURN	36	37	
FT	HELIX	40	44	
FT	TURN	45	46	
FT	STRAND	47	47	
FT	TURN	48	50	
FT	STRAND	51	51	
FT	HELIX	62	78	
FT	TURN	79	79	
FT	HELIX	81	84	
FT	TURN	85	86	
FT	TURN	90	92	
FT	STRAND	97	100	
FT	TURN	101	102	
FT	STRAND	103	106	
FT	TURN	108	109	
FT	HELIX	112	130	
FT	HELIX	135	137	
FT	STRAND	138	138	
FT	TURN	139	139	
FT	HELIX	142	143	
EQ	SEQUENCE	146 AA;	16279 MM;	DE87674C9476FA36 CRC64;
Alignment Scores:				
Pred. No.:	5.63	Length:	146	
Score:	74.50	Matches:	29	
Percent Similarity:	33.06%	Conservative:	11	
Best Local Similarity:	23.97%	Mismatches:	44	
Query Match:	11.05%	Indels:	37	
BB:	1	Gaps:	4	

	US-09_-936-1737A-1 (1-375) x PAZL_PIG (1-146)	
QY	16 ATTTCTTCCGTCCTGGCGCAAGCTGTGATCTCAACTACTCTTCAGAA-----GAA	69
Dd	1 MetLVSPheLeuValLeuAlaValLeuThrValcIyAlaLaIngluNIyleSer	20
QY	70 CGTGAAGATTGTTGACGCTTTTACGCAGACAGAAA-----	105
Dd	21 SerArgAlaLeuTrpGIlnPheArgSerMetIleYscysAlaIleProIlyserHisPro	40
QY	106 -----TANRACAGACTTGCAATAATCCTTTTAAGAAGTCC	138
Dd	41 LeuMetAspPheAsnAsnTrgIylCYSTrYCysGlYleuGlYglYSerGIlythrProVal	60
QY	139 TCGATCTTCGACGAATGGAAAAAACMTGTTTCAAGCGSAGTACGTCNACTGTTTT	198
Dd	61 AspGIlnLeuAsPArGYcysGIlnthr-----	69
QY	199 GAAGACACGCTCACACAAGAAATTACTACAAATGCTGTGATGCGTGAACAGTTTAGCCA	258
Dd	70 -----HisAspAsnCysTyArg-----AspAlaIlyasmsneuAsper	82
QY	259 GAAAAATTGTTGTGTCGACAGAAACTTCACGGAATATTATTGACAGACTGCCAGSGTAA	318
Dd	83 CysLVSPheLeuValAsPAsnPTrOlyThrlusertYlserTYlserCYScysSeasnThr	102
QY	319 GAT 321	
Dd	103 Glu 103	
RESULT 13		
Z222_HUMAN		
ID	Z222_HUMAN	STANDARD;
AC	O9UK12;	PRT; 451 AA.
DT	16-OCT-2001 (Rel. 40, Created)	
DT	16-OCT-2001 (Rel. 40, Last sequence update)	
DT	16-OCT-2001 (Rel. 40, Last annotation update)	
DE	Zinc finger protein 222.	
DN	ZNF222.	

OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RA Shannon M., Branscomb E., Hauser L., Gordon L., Ashworth L.,
 RA Stubbs L.;
 RT "Differential expansion of homologous zinc-finger gene families in
 human chromosome 19q13.2 and mouse chromosome 7.",
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
 CC FINGER PROTEINS.
 CC -1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF187988; AAF04104.1; -.
 DR HSSP: P03001; JTF6
 DR GeneW: HGNC:13015; ZNF222.
 DR InterPro: IPR001909; KRAB.
 DR InterPro: IPR000822; Znf_C2H2.
 DR Pfam: PF00096; Zf_C2H2; 8.
 DR Pfam: PF01352; KRAB; 1.
 DR ProDom: PD000003; Znf_C2H2; 5.
 DR SMART: SM00349; KRAB; 1.
 DR SMART: SM00355; Znf_C2H2; 8.
 DR PROSITE: PS50805; KRAB; 1.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 8.
 DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 10.
 KW Transcription regulation; DNA-binding; zinc-finger; Metal-binding;
 KW Nuclear protein; Repeat.
 FT DOMAIN 8
 FT DOMAIN 85
 FT ZN_FING 143 391 ZINC FINGERS.
 FT ZN_FING 145 167 C2H2-TYPE.
 FT ZN_FING 173 195 C2H2-TYPE.
 FT ZN_FING 201 223 C2H2-TYPE.
 FT ZN_FING 229 251 C2H2-TYPE.
 FT ZN_FING 257 279 C2H2-TYPE.
 FT ZN_FING 313 335 C2H2-TYPE.
 FT ZN_FING 341 363 C2H2-TYPE.
 FT ZN_FING 369 391 C2H2-TYPE.
 SQ SEQUENCE 451 AA; 52532 MW; 55F55A778116D184 CRC64;
 Alignment Scores:
 Pred. No.: 6.05 Length: 451
 Score: 74.50 Matches: 33
 Percent Similarity: 40.91% Conservative: 21
 Best Local Similarity: 25.00% Mismatches: 43
 Query Match: 11.05% Indels: 35
 DB: 1 Gaps: 6
 US-09-936-737a-1 (1-375) x Z222.HUMAN (1-451)
 QY 46 ATCTCACTACTTCTTCAAGAAAGCT-----GAAGATTGTTGAGCGTTTTCAC 93
 DB 107 LeuSerThrValHisThrArgGluLysProPheGlnGluAsnGlyGlnPheSer 126
 QY 94 GCGAAGCAAAATATACAGACTTCGATAA-----TCTTTAAGAACTCCTGTAT 144
 DB 127 SerAspValSerPhePheAspLeuProGlnGlnLeuTyrSerGlyGluLysSerHisThr 146
 QY 145 CTTCAGCAATGCAAAAAACA----- 165
 DB 147 CysAspGlnCysGlyLysSerPheCysTyrIleSerAlaLeuHisIleHisGlnArgVal 166

QY 166 -----TGTTCAGACGAGTACTGTCATCGTTTTCAGACAGGTC 210
 DB 167 HisMetGlyValLysCysTyrTyrCysAspValCysGlyLysGluPheSerIleSer 186
 QY 211 AACAGCAATGTTACTACATGTCGTTGATGCGAAGC---TTAGACCAAGAAAAATT 267
 DB 187 ArgLeuGlnThrHisGlnArgValHisThrGlnGluLysProPheLysCysGlnGln--C 206
 QY 268 GT-----TGTCCAGCA-----AACTTACGGAATATATTTCACAGC 306
 DB 206 yslGlyLysGlyPheArgCysArgSerAlaLeuLysValHisCysLysIleHisMetArg 226
 QY 307 TCCGAGGCTAAAGATGATGATGCGCAGG 338
 DB 226 IuLysProTyrAsnCysGluLysCysGlyLys 236
 RESULT 14
 ID YJ23_AQUAE STANDARD; PRT; 296 AA.
 AC 067752;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein AQ_1923 precursor.
 GN AQ_1923.
 OS Aquifex aeolicus.
 OC Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;
 OC Aquifex.
 OX NCBI_TaxID=63363;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VF5;
 RX MEDLINE=98196666; PubMed=9537320;
 RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
 RA Feldman D.E., Overbeek R., Sneed M.A., Keller M., Anjey M., Huber R.,
 RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
 RT "The complete genome of the hyperthermophilic bacterium Aquifex
 RT aeolicus".
 RL Nature 392:353-358(1998).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AEO00763; AAC07719.1; -.
 KW Hypothetical protein; Signal; Complete proteome.
 KW POTENTIAL.
 FT CHAIN 1 296
 FT SIGNAL 1 20
 SQ SEQUENCE 296 AA; 34566 MW; 9C88F456AA9449C CRC64;
 Alignment Scores:
 Pred. No.: 6.64 Length: 296
 Score: 74.00 Matches: 22
 Percent Similarity: 43.96% Conservative: 18
 Best Local Similarity: 24.18% Mismatches: 21
 Query Match: 10.98% Indels: 30
 DB: 1 Gaps: 4
 US-09-936-737a-1 (1-375) x YJ23_AQUAE (1-296)
 QY 88 TTTTACGCGAAGCAAAATATACAGC-----TTC 117
 DB 18 TyrPheIleTyrAspLysTyrMetAspAsnLysAlaLysGluGlnValGluTyrPheLeu 37
 QY 118 GATTAATCTTTTAAAGAACTCCTGTATCTTGAGCAATGCAAAAAAACATGTTTCAGAGC 177
 DB 38 AspLysThrLeuArgLysSerGly-----LysGly 47

```

OY 178 GAGTACTGCTACATCGTTTGA-----GACACGGTCAACAAGAAATGTAC 225
    ||| |||::: ::::: |||::: |||::: |||
DB 48 SerTylTyrValaspIrrLysProIleGlyGluIleIleLysaspValTyr 67
    ||| |||::: |||::: |||::: |||::: |||
OY 226 TACATGTCGTGTAGTGAAGATTAGACCAAGAAAAATTTGTGTCGACGAAAACTTC 285
    ||| |||::: |||::: |||::: |||::: |||
DB 68 TYARqAspArgasnGlyGluGluPheLysIleGluGluIleIleIleGluLys----- 85
    ||| |||::: |||::: |||::: |||::: |||
OY 286 ACGGAATAATTATTTGACACACCTGCCGAGGTAAA 318
    ||| |||::: |||::: |||::: |||::: |||
DB 86 -----LeuSerGluThrGluGlyLys 92
    ||| |||::: |||::: |||::: |||::: |||

RESULT 15
SP96.DICDI
ID SP96.DICDI STANDARD; PRT; 600 AA.
AC P14328;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Spore coat protein SP96.
GN COXA
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
ON NCBI_TaxID=4689;
RX MEDLINE=90067962; PubMed=2587278;
RA Fornaugh K., Loomis W.F.;
RL Nucleic Acids Res. 17:9489-9489(1989).
CC -1 SUBCELLULAR LOCATION: OUTER LAYER OF THE COAT MATRIX AND
    INTERSPORE MATRIX.
CC -1 PPM: PHOSPHORYLATED AND FUCCOSYLATED. MAY BE PHOSPHOGLYCOSYLATED,
    MAY CONTAIN GLCNAC-ALPHA-1-P-SER RESIDUES.
CC -1 SIMILARITY: CONTAINS 4 PRESPORE MOTIFS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: X16491; CAA34508.1; -
DR PIR: S07638; S07638.
DR DR GlycoSuiteDB; P14328; -
DR DictyDB; DD03007; COXA.
DR InterPro: IPR003645; FOLN.
DR SMART: SM00274; FOLN; 6.
KW Glycoprotein; Phosphorylation; Repeat; Sporulation.
FT REPEAT 185 197
FT REPEAT 221 233 PRESPORE MOTIF 1.
FT REPEAT 298 310 PRESPORE MOTIF 2.
FT REPEAT 395 407 PRESPORE MOTIF 3.
FT REPEAT 407 PRESPORE MOTIF 4.
SQ SEQUENCE 600 AA; 59589 MW; 616AE6D02B5F1071 CRC64;

Alignment Scores:
Pred. No.: 6.95 Length: 600
Score: 74.00 Matches: 21
Percent Similarity: 58.70% Conservative: 6
Best Local Similarity: 45.65% Mismatches: 19
Query Match: 11.37% Indels: 0
DB: 1 Gaps: 0

US-09-936-737a-1 (1-375) x SP96.DICDI (1-600)
OY 371 TCTTCAATCAACTGATCTGATCACTGTACCTGCCGATTTACTGCATCTTTACCC 312
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 488 SerSerSerAlaSerSerSerProSerSerSerAlaSerSerSerAlaPro 507
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 311 TCGCAGTCGTCAATAATTTTCCGTGAAGTTTCTGTCGACCAACAATTTTCTTGCT 252
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

DB 508 SerSerSerAlaSerSerSerSerAlaProSerSerSerAlaSerSerSerAlaSer 527
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 251 AACTCTTCACCATCAACG 234
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 528 SerSerSerAlaSerSer 533
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Search completed: February 25, 2003, 02:26:31
Job time : 16.5 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 25, 2003, 01:54:30 ; Search time 44 Seconds
(without alignments)
3512.167 Million cell updates/sec

Title: US-09-936-737A-1
Perfect score: 674
Sequence: 1 atgaagtattcttgattc.....atgaagtatgaatgaatgaat 375

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 671580 segs, 206047115 residues
Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+nrp.model -DEV=xlh
-Q=/cgn2.1/USPTO.spool/US09936737/runat.14022003.100557.1398/app-query.fasta.1.519
-DB=SPREMBL.21 -QEXT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPEXT=0 -LIST=45
-UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pro -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000 -ICPU=3
-USRR=US09936737 -GCN1.1.34 -runat.14022003.100557.1398 -NCPU=6 -ICPU=3
-NO_XLRY -NO_MMAP -LARGEQUERY -NEG.SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPREMBL.21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organella:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeap:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	89.5	13.3	3933	5 097239 plasmodium

2	84.5	12.5	164	5	044835	044835 caenorhabdi
3	82.5	12.2	368	5	09vfh8	09vfh8 drosophila
4	82.5	12.2	374	5	08sxc5	08sxc5 drosophila
5	82	12.6	637	10	080625	080625 arabidopsis
6	82	12.2	907	5	09xts9	09xts9 caenorhabdi
7	81	12.0	388	16	08r7b3	08r7b3 thermoaer
8	80.5	11.9	1303	10	09sk50	09sk50 arabidopsis
9	79.5	12.2	550	10	094by9	094by9 arabidopsis
10	79.5	11.8	1277	11	035821	035821 rattus norv
11	79	11.7	358	5	08sra3	08sra3 encephalito
12	78.5	11.6	116	16	08xj41	08xj41 clostridium
13	78.5	11.6	264	5	020246	020246 caenorhabdi
14	78.5	11.6	509	2	09xys1	09xys1 streptococ
15	78.5	11.6	550	5	09u204	09u204 caenorhabdi
16	78	11.6	682	16	09k8n5	09k8n5 bacillus ha
17	77.5	11.5	191	5	02052	02052 caenorhabdi
18	77	11.4	257	5	09m557	09m557 caenorhabdi
19	77	11.4	355	16	09sp06	09sp06 mycoplasma
20	77	11.4	945	5	003650	003650 trypanosoma
21	76.5	11.4	168	10	09lq13	09lq13 arabidopsis
22	76.5	11.4	343	8	08w921	08w921 demansia at
23	76.5	11.4	786	10	09c7e7	09c7e7 arabidopsis
24	76	11.3	1255	16	097116	097116 clostridium
25	76	11.3	1278	16	050667	050667 borrelia bu
26	75.5	11.2	336	10	09s200	09s200 arabidopsis
27	75.5	11.2	509	8	09tkr7	09tkr7 lindenbergt
28	75.5	11.2	703	16	099263	099263 streptococ
29	75.5	11.2	719	16	099263	099263 streptococ
30	75	11.2	210	3	09us87	09us87 schizosacch
31	75	11.1	324	2	09s0m3	09s0m3 staphylococ
32	75	11.5	344	3	09s109	09s109 schizosacch
33	75	11.5	534	3	094317	094317 schizosacch
34	75	11.1	782	10	09lkr5	09lkr5 arabidopsis
35	75	11.1	2410	3	09lkr5	09lkr5 schizosacch
36	75	11.1	3844	5	094648	094648 plasmodium
37	74.5	11.1	146	10	08s554	08s554 rosa hybrid
38	74.5	11.1	156	11	09c049	09c049 mus musculu
39	74.5	11.1	328	3	012140	012140 saccharomyc
40	74.5	11.1	451	4	09p105	09p105 homo sapien
41	74.5	11.1	469	2	08vwm8	08vwm8 listeria se
42	74.5	11.1	606	10	09f1w8	09f1w8 arabidopsis
43	74.5	11.0	643	5	09xv53	09xv53 caenorhabdi
44	74	11.0	248	5	09vrf9	09vrf9 drosophila
45	74	11.0	292	12	065239	065239 african swi

ALIGNMENTS

RESULT 1
ID 097239 PRELIMINARY; PRT; 3933 AA.

AC 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Hypothetical 467.9 kDa protein.
OS PFC0245C, MAL3P2.18.
GN Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=99376085; PubMed=10448855;
RA Bowman S.C., Lawson D., Basham D., Brown D., Chillingworth T.,
RA Churcher C.M., Craig A., Davies R.M., Devlin K., Feltwell T.,
RA Gentles S., Gilliam R., Hamlin N., Harris D., Holroyd S., Hornsby T.,
RA Horrocks P., Jagels K., Jassal B., Kyes S., McLean J., Moulie S.,
RA Mungall K., Murphy L., Oliver K., Quail M.A., Rajandream M.-A.,
RA Rutter S., Skelton J., Squares R., Squares S., Sulston J.E.,
RA Whitehead S., Woodward J.R., Newbold C., Barrett B.G.,
RT "The complete nucleotide sequence of chromosome 3 of Plasmodium
falciparum .";

Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfei

RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA April J.F., Agdayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
 RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durdin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasner K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kenison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palzozo L.M., Piltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Relbert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spler E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskaes R., Tector C., Turner R., Venter G.M., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Glibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of *Drosophila melanogaster*."
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003705; AARS5078.1;
 DR FlyBase: FBgn0058246; CG14853.
 SQ SEQUENCE 368 AA; 38008 MW; 298927508C12DBFC CRC64;

Alignment Scores:

Pred. No.: 3 25 Length: 368
 Score: 82.50 Matches: 34
 Percent Similarity: 42.14% Conservative: 25
 Best Local Similarity: 24.29% Mismatches: 42
 Query Match: 12.24% Indels: 39
 DB: 5 Gaps: 7

US-09-936-737A-1 (1-375) x Q9VFNH (1-368)

QY 34 GCAAGCTTGGTATCTCACTACTTCTTCAGAGAAGCTGAAGATTGTTGACGTTTAC 93
 DB 71 AAlserGlyLeuProThrSerAlaSerSerGluAspLeuSerGlnSerLeuSer----- 88
 QY 94 GCGAAGCAAGAAATATACAGACTTCGATTAATCT----- 126
 DB 89 -----GluThrThrAspAlaAspGluSerValSerAlaProThrGluPheLeuAla 105
 QY 127 ---TTTAAGAAGCTCTCTGATCTTTCAGCAATGCAAAAACATGTTTCAAGCGAGTAC 183
 DB 106 GluPheLeuSerAlaValMetLeuLysAspTyrLysVala-----LeuLysTyr 122
 QY 184 TGCTACATCGTTTGTGA-----GACACGGTCAACAGGAATGTTACTACAAATGTC 234
 DB 123 CysLysLeuLeuLeuGlnTyrGluProAspAsnAlaThrAlaLysGluPheTyrProLeu 142
 QY 235 GTTGATGTGAAGAGTTTACCAAGAAAATTTGTTGTC-----GAGGAA 279
 DB 143 Ile-----LeuAspLysLeuArgAlaValAlaThrSerSerAspSerAspGlu 158
 QY 280 AACTTCACGGAAGAAAT-----TATTGACAGACTGCGAG 312
 DB 159 AsnTyrAsnLysSerSerSerProAspLeuAlaLeuAspLeuHisAlaSerAspValGlu 178
 QY 313 GGTAAAGATGACGATATGCGGACGATACAGTGCAGCGATGAGTGAAGTGAAGAT 372
 DB 313 GGTAAAGATGACGATATGCGGACGATACAGTGCAGCGATGAGTGAAGTGAAGAT 372

DB 179 AlaAspValAspGlyAspGluAlaGlyAspAlaAspGluAspGlyAspAlaAspAlaAsp 198
 RESULT 4
 ID 08SYG5 PRELIMINARY: PRT: 374 AA.
 ID 08SYG5
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE RE63320P.
 GN CG14853.
 OS *Drosophila melanogaster* (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachytera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NCBI_TaxID=7227;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY.
 RA Stapleton M., Brokstein P., Hong L., Agdayani A., Carlson J.,
 RA Champagne M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
 RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,
 RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
 RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
 RA Gelniker S.,
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY071572; AL49194.1;
 SQ SEQUENCE 374 AA; 38729 MW; 914936D5E40AF93A CRC64;

Alignment Scores:

Pred. No.: 3 24 Length: 374
 Score: 82.50 Matches: 34
 Percent Similarity: 42.14% Conservative: 25
 Best Local Similarity: 24.29% Mismatches: 42
 Query Match: 12.24% Indels: 39
 DB: 5 Gaps: 7

US-09-936-737A-1 (1-375) x Q8SYG5 (1-374)

QY 34 GCAAGCTTGGTATCTCACTACTTCTTCAGAGAAGCTGAAGATTGTTGACGTTTAC 93
 DB 77 AAlserGlyLeuProThrSerAlaSerSerGluAspLeuSerGlnSerLeuSer----- 94
 QY 94 GCGAAGCAAGAAATATACAGACTTCGATTAATCT----- 126
 DB 95 -----GluThrThrAspAlaAspGluSerValSerAlaProThrGluPheLeuAla 111
 QY 127 ---TTTAAGAAGCTCTCTGATCTTTCAGCAATGCAAAAACATGTTTCAAGCGAGTAC 183
 DB 112 GluPheLeuSerAlaValMetLeuLysAspTyrLysVala-----LeuLysTyr 128
 QY 184 TGCTACATCGTTTGTGA-----GACACGGTCAACAGGAATGTTACTACAAATGTC 234
 DB 129 CysLysLeuLeuLeuGlnTyrGluProAspAsnAlaThrAlaLysGluPheTyrProLeu 148
 QY 235 GTTGATGTGAAGAGTTTACCAAGAAAATTTGTTGTC-----GAGGAA 279
 DB 149 Ile-----LeuAspLysLeuArgAlaValAlaThrSerSerAspSerAspGlu 164
 QY 280 AACTTCACGGAAGAAAT-----TATTGACAGACTGCGAG 312
 DB 165 AsnTyrAsnLysSerSerSerProAspLeuAlaLeuAspLeuHisAlaSerAspValGlu 184
 QY 313 GGTAAAGATGACGATATGCGGACGATACAGTGCAGCGATGAGTGAAGTGAAGAT 372
 DB 185 AlaAspValAspGlyAspGluAlaGlyAspAlaAspGluAspGlyAspAlaAspAlaAsp 204
 RESULT 5
 ID 080625 PRELIMINARY: PRT: 637 AA.
 ID 080625
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)

DT 01-DEC-2001 (T-EMBLrel. 19, last annotation update)
 DE ATG3380 protein.
 GN ATG3380.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;
 CC Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae;
 OC Eucosids II: Brassicales; Brassicaceae; Arabidopsis.
 NC NCB1_TaxID=3702;
 RX MEDLINE=20083487; PubMed=10617197;
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
 RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
 RA Cronin L.A., Shen M., Vanaken S.E., Unayam L., Tallon L.J., Gill J.E.,
 RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
 RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
 RA Salzberg S.L., Fraser C.M., Venter J.C.;
 RA "Sequence and analysis of chromosome 2 of the plant Arabidopsis
 thaliana";
 RL Nature 402:761-768(1999).
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Lin X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC004218; AAC27829.1;
 DR InterPro: IPR004140; Exo70.
 DR Pfam: PF03081; Exo70; 1.
 SQ SEQUENCE 637 AA; 72179 MW; 03C4606CBB8335F CRC64;
 Alignment Scores:
 Pred. No.: 3.61 Length: 637
 Score: 82.00 Matches: 38
 Percent Similarity: 48.91% Conservative: 29
 Best Local Similarity: 27.74% Mismatches: 30
 Query Match: 12.60% Indels: 40
 DB: 10 Gaps: 9
 US-09-936-737a-1 (1-375) x 080625 (1-637)
 QY 365 TCACATTCATCTGACGTCGCA-----CCGTGACCTGCCGACATCTACGCA 321
 DB 11 SerThSerSerProLysSerPheLeuSerSerPheProSerPheThSerLeuProLa 30
 QY 320 TCCTTACCC-----TCGACGCTGTCATAATATTTTCGCGAAGTTTCG 276
 DB 31 Ser---ProLeuAnGlnThrPheSerGlnSerMetMetGluGluThValGluAlaLa 49
 QY 275 TCGACAAACAATTTTCTTGCTTAACCTTCACATCAGACACATGTGAGTAACATTC 216
 DB 50 GluSerIleIleLysLysTrpAspProAnSerProSer----- 62
 QY 215 TTGTGACCGTGTTCACAAAGCATGTAGACTGCTGCTGTAACATGTTTTTTTG 156
 DB 63 -----TyrThLysIleIleSerLeuHeSer 71
 QY 155 CATTCGTCAGAG--TCAGAGAGATTC-----TTAAAGATTTCAGAGCTGTGA 108
 DB 72 HisSerTrgArGluAlaLysGluPheIleArgCysIleArgspLeuTrgArGluMet 91
 QY 107 TATTTCTGTCGCGGTAACAGCTTCACAAATCTTCACGCTTCCTTCAGACAGAGT 51
 DB 92 HisPheLeu-----IleSerGlnHisSerLysSerAlaLysLeuValLeuAla 107
 QY 50 ---GAGATCAAGCAAGCTTCGAGAGCAAGCAAGCAAGCAAGCAAGCAAGTCTTC 3
 DB 108 GlnHisLeuMetGlnIleAlaMetAlaArgLeuGlu---LysGluPhePhe 123
 RESULT 6
 Q9XTS9 PRELIMINARY; PRT; 907 AA.

AC Q9XTS9;
 DT 01-NOV-1999 (T-EMBLrel. 12, Created)
 DT 01-NOV-1999 (T-EMBLrel. 12, last sequence update)
 DT 01-MAR-2002 (T-EMBLrel. 20, last annotation update)
 DE Y69H2.2 protein.
 GN Y69H2.2.
 OS Caenorhabditis elegans.
 OC Eukaryota: Metazoa: Nematoda: Chromadorea: Rhabditida: Rhabditidae;
 CC Rhabditidae; Peloderinae; Caenorhabditis.
 NC NCB1_TaxID=6239;
 RX MEDLINE=990659613; PubMed=9851916;
 RA none;
 RL "Genome sequence of the nematode C. elegans: A platform for
 investigating biology";
 RL Science 282:2012-2018(1998).
 DR EMBL: Z98877; CAB1570.1; -.
 DR HSSP; Q12784; HRE. EGF-like.
 DR InterPro: IPR000561; EGF-1.
 DR InterPro: IPR001438; EGF-II.
 DR InterPro: IPR001873; Na+channel_ASC.
 DR Pfam: PF00088; ASC; 1.
 DR Pfam: PF00008; EGF; 7.
 DR PRINTS: PRO1078; AMINCHANNEL.
 DR PRINTS: PRO0010; EGFBLD.
 DR SMART; SM00181; EGF; 8.
 DR PROSITE; PS00022; EGF_1; UNKNOWN_8.
 DR PROSITE; PS01186; EGF_2; 4.
 KW EGF-like domain; Glycoprotein.
 SQ SEQUENCE 907 AA; 101840 MW; 8B55FF139F8F81D7 CRC64;
 Alignment Scores:
 Pred. No.: 3.57 Length: 907
 Score: 82.00 Matches: 37
 Percent Similarity: 39.85% Conservative: 16
 Best Local Similarity: 27.82% Mismatches: 40
 Query Match: 12.17% Indels: 40
 DB: 5 Gaps: 9
 US-09-936-737a-1 (1-375) x Q9XTS9 (1-907)
 QY 46 ATCTCACTACTCTTCACAGAGACGTAAGATTGTTGACGTTT----- 90
 DB 206 IleGlyThHisGlnThrLysTrpLysAspAspLysMetTrpSerAspLysCysAnGlu 225
 QY 91 TACGCGAAGAAATATATATACAGACTTCGATTAATCTTTAA----- 132
 DB 226 TyrAlaGlnGluLysSerThrAsp---AspLysThrLeuLysTrpLeuThrLeuCysGly 244
 QY 133 -----AAGTCCTGATCTT-----GACGA 153
 DB 245 GluTrpCysMetValSerGluLysPheLeuValLysAspAspValSerMetLysAsp 264
 QY 154 TGCAGAAAGACATG-----TTCAAGAGGAGTACTGCTACATCGTTTGAAGACACG 207
 DB 265 CysLysLysSerCysGlyLysPheSerSerGluTrpCys-----GlyLys 279
 QY 208 GTCAACAGAGATGTACTACATGCTGCTGATGCGAGAGTACAGACGAAATTT 267
 DB 280 ValAsnSerArgCysIle-----ValTyrGluAlaGlyGluMetGlu----- 293
 QY 268 GTTGTACAGAAACTTCACGAGAAATTTATTTGACAGACTGCGAGGTAAGATCAGCT 327
 DB 294 ---ThrAspProAnAlaCysGluAspAsnSerThrLeuCys---GlyAlaGluLeuGly 311
 QY 328 AATCGCGAGGTACAGGTGACGATGATGAAGTTGAT 366
 DB 312 HisGlyMetCysIleAsnTrpGlnSerAspValThrAsp 324

RA Katinka M. V., Dupluc C., Brothier P., Winckel F.,
RA Prensier G., Barbe V., Peyretallade E.,

RA Delbac F., El Alaoui H., Peyret P., Saurin W., Gouy M.,
 RA Weissensbach J., Vayres C.P.,
 RT "Genome sequence and gene compaction of the eukaryote parasite
 RL Encephalitozoon cuniculi".
 DR Nature 414:450-453(2001).
 S0 EMBL: AL590448; CND26435.1; -
 SEQUENCE 358 AA; 40074 MW; CCABFD0245F98342 CRC64;

Alignment Scores:

Pred. No.: 7.57 Length: 358
 Score: 79.00 Matches: 28
 Percent Similarity: 41.74% Conservative: 20
 Best Local Similarity: 24.35% Mismatches: 41
 Query Match: 11.72% Indels: 26
 DB: 5 Gaps: 5

US-09-936-737a-1 (1-375) x Q8SRA3 (1-358)

QY 22 TTCCTTGGCTGCGACGCTG---CTGATCTCACTACTTCTTGAAGAAGTGAAGAT 78
 Db 239 PheCysCysLeuSerThrLeuArgSerLeuGlyAlaLeuGluThrValAspArgGluGln 258
 QY 79 TGTGGACGTTTACGCGAAGAAATATACAGACTTCGATAATCTTTAAGAAATCC 138
 Db 259 ValAlaArgPheLeuAlaThrLysGlnAlaSerSerGlyLeuSerGlyValSer 278
 QY 139 TCTGATCTTGACGATGCAATGCAAAAAAACATGTTTCAAGACGAGTACTGCTACATCGTTT 198
 Db 279 LysLysGluAspValCys-----TyrSerPheTrpAlaLys 290
 QY 199 GAAGACACGCTC-----AACAGAAATGTTACTACAGTCGTTGATGTTGAAGATTA 252
 Db 291 SerSerLeuValLeuIleGlyLysGluCysTyr-----ValAsnGlnGluGln 307
 QY 253 GACCAAGAAAAATTTGTTGTCGACGAAACTCACGGAATATTTTACAGACTGCGAG 312
 Db 308 ThrArg-----PheLysPheSerCysGln 315
 QY 313 GGTAAAGATGCGAGTAATGCGCGAGGTACAGGTGACGAGTCAGAT 357
 Db 316 GlyProSerGlyGlyPheSerAspArgProGlyAsnGluThrAsp 330
 RESULT 12

ID 08XJ41 PRELIMINARY; PRT; 116 AA.
 AC 08XJ41;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 OS CPEI1920.
 ON Clostridium perfringens.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
 OX NCBI_Taxid=1502;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN=13 / TYPE A;
 RX PUBMED=11792842;
 RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
 RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.,
 RT "Complete genome sequence of Clostridium perfringens, an anaerobic
 RL flesh-eater".
 RL EMBL: AP003192; BAB81626.1; -
 DR InterPro: IPR001440; TPR.
 DR Pfam: PF00515; TPR; 3.
 DR SMART: SM00028; TPR; 3.
 KW Hypothetical protein; Complete proteome.
 S0 SEQUENCE 116 AA; 13590 MW; 9DB58D847B3A7312 CRC64;

Alignment Scores:

Pred. No.: 8.81 Length: 116

Score: 78.50 Matches: 22
 Percent Similarity: 34.02% Conservative: 11
 Best Local Similarity: 22.68% Mismatches: 27
 Query Match: 11.65% Indels: 37
 DB: 16 Gaps: 2

US-09-936-737a-1 (1-375) x Q8XJ41 (1-116)

QY 91 TACGCGAAGAAATATATACAGCTTGCATTAATCTTTAAGAACTCTGATCTTGAC 150
 Db 11 TyrAsnThrArgAsnTyrSerAspAlaIleAsnTyrTyrLysAlaLeuAspAsp 150
 QY 151 GAATGCAAA----- 159
 Db 31 GlucylsCysHisSerTyrTyrAsnAlaGlyValCysTyrIleLysLeuLysGlnTyr 159
 QY 160 ----- 168
 Db 51 GluLysAlaIleGluMetIleThrLysAlaLeuGluLeuTyrGlnAspSerLysTyrPhe 168
 QY 169 TTAAGACGAGACTGCTACTACGTTTGTGAAGACAGGTCACAAAGAAATGTTACTAC 228
 Db 71 PheAsnLeuAlaTyrCysTyrSerMetIleAsnAsnAsnSerLysAlaLeuArgTyrPhe 228
 QY 229 ATATCT-----GTTGATGTTGAAGACTTGAACCAAGAAATTTGTT 270
 Db 91 AsnLeuAlaTyrPalaLeuAspAsnAlaAspIleAspCysGluLysAlaIle 107
 RESULT 13

ID 020246 PRELIMINARY; PRT; 264 AA.
 AC 020246;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 GN F40G12.5 protein.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Pelodermidae; Caenorhabditis.
 OX NCBI_Taxid=6239;
 RN 11
 RP SEQUENCE FROM N.A.
 RC McMurtry A.A.;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 RN 12
 RP MEDLINE=95069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C.elegans: A platform for
 RL investigating biology".
 RL Science 282:2012-2018(1998).
 DR EMBL: Z77661; CAB01183.1; -
 DR InterPro: IPR004920; DUF263.
 DR Pfam: PF03236; DUF263; 2.
 S0 SEQUENCE 264 AA; 30508 MW; FD237BB570436204 CRC64;

Alignment Scores:

Pred. No.: 8.61 Length: 264
 Score: 78.50 Matches: 35
 Percent Similarity: 41.54% Conservative: 19
 Best Local Similarity: 26.92% Mismatches: 47
 Query Match: 11.65% Indels: 29
 DB: 5 Gaps: 8

US-09-936-737a-1 (1-375) x Q20246 (1-264)

QY 10 TTTGATTTGCTCTGCTTGGCTGCGAAGCTTGCATCTCACTACTTCTTGAAGAA 69
 Db 11 PheAlaIleSerAlaAsnCysSerIleLeuSerPheHisArgAsnLeuGluGluGln 30
 QY 70 CGTGAAGATTGTTGACGTTTACGCGAAGAAATATACAGACTTCGATAATCTTTT 129
 |||||

[illegible][illegible]

Search completed: February 25, 2003, 02:28:15
job time : 50 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2003, 00:58:30 ; Search time 1963 seconds
(without alignments)
3093.891 Million cell updates/sec

Title: US-09-936-737A-1

Perfect score: 375
Sequence: 1 atgaagtattcttgattc.....atgaagtgtgaagattaa 375

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	51.8	13.8	879	AZ550718 ENT50718
2	51.6	13.8	701	BM168486 EST571009
3	51.6	13.8	721	BM165474 EST567997
4	51.2	13.7	881	AZ688514 ENT5022TF
5	50	13.3	236	BM163498 EST566021
6	50	13.3	272	BM166089 EST568612

7	50	13.3	274	13	BM166886	EST569409
8	50	13.3	276	13	BM163293	EST565816
9	50	13.3	425	13	BM162507	EST565030
10	50	13.3	453	13	BM167376	EST566899
11	50	13.3	459	13	BM171330	EST573853
12	50	13.3	580	13	BM161175	EST563698
13	50	13.3	648	13	BM164042	EST565655
14	50	13.3	651	13	BM166448	EST568971
15	50	13.3	667	13	BM161293	EST563816
16	50	13.3	678	13	BM170834	EST573357
17	50	13.3	702	13	BM169291	EST571814
18	50	13.3	719	13	BM602689	EST501779
19	50	13.3	739	13	BM160442	EST562965
20	50	13.3	753	13	BM167115	EST569638
21	50	13.3	762	12	BE846498	PSB 206 L
22	50	13.3	790	13	BM166415	EST568938
23	49.2	13.1	942	17	BM148582	ENTPR62TF
24	49	13.1	816	17	AZ535744	ENTCQ25TF
25	49	13.1	891	17	AZ683582	ENTRK47TF
26	49	13.1	976	17	BM149983	ENTOD93TF
27	48.4	12.9	435	13	BM169710	EST572233
28	47.4	12.6	900	17	AZ549980	ENTDP94TF
29	47	12.5	843	17	AZ551618	ENTDV54TF
30	47	12.5	877	17	AZ531291	ENTBQ34TF
31	47	12.5	908	17	AZ548467	ENTEK30TF
32	47	12.5	912	17	AZ551092	ENTFJ22TF
33	47	12.5	931	17	BM160272	ENTOV49TF
34	46.8	12.5	724	13	BM161274	EST563797
35	46.8	12.5	1006	17	AZ679713	ENTH56TF
36	46.6	12.4	558	13	BM163267	EST565790
37	46.6	12.4	890	17	BM146886	ENTPK48TF
38	46.6	12.4	906	17	BM153606	ENTTS83TF
39	46.4	12.4	468	12	BG553005	dab82e09.
40	46.4	12.4	880	17	AZ669474	ENTY88TF
41	46.2	12.3	493	13	BM170559	EST573082
42	46.2	12.3	1101	17	AL108773	Drosophila
43	45.8	12.2	849	17	AZ546009	ENTFW3TF
44	45.8	12.2	905	17	AZ550256	ENTEW58TF
45	45.4	12.1	916	17	AZ671886	ENTLR44TF

ALIGNMENTS

RESULT 1
AZ550718/c 879 bp DNA linear GSS 14-NOV-2000
LOCUS ENT50718 Entamoeba histolytica sheared DNA Entamoeba histolytica
DEFINITION ENT50718 genomic DNA sequence.
ACCESSION AZ550718 GI:11176019
VERSION GSS.
KEYWORDS Entamoeba histolytica.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica.
REFERENCE 1 (bases 1 to 879)
Loftus, B., Van Aken, S. and Fraser, C.
Determination of clone end sequences from Entamoeba histolytica
Eukaryota; Entamoebidae; Entamoeba.
TITLE HMI:IMSS sheared DNA library
JOURNAL Unpublished (2000)
COMMENT Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@igf.org

Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library
Seq primer: M13-Forward
Class: Shotgun
High quality sequence start: 25
High quality sequence stop: 801.

FEATURES

Location/Qualifiers
1..879

Source

/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pHS1; Site 1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999)."

BASE COUNT

193 a 219 c 72 g 395 t

ORIGIN

Query Match

13.8%; Score 51.8; DB 17; Length 879;

Best Local Similarity

46.9%; Pred. No. 0.0004;

Matches 161; Conservative

0; Mismatches 182; Indels 0; Gaps 0;

QY 29

GCCTGCAAGCTTCTGCTCAACTCTCTCAGAGACGGAAGATTGTGGACGT 88

DB 768

GACTGTGTTACTGCTTTAATGCTTTGCCGAGAAATGAACTTGACACACAT 709

QY 89

TTTACCGCAACAGAAATATACAGACTTCGATTAATCTTTTAAAGTCTGTATCTG 148

DB 708

TGCAATTTGATTTCTACGTAGTATACGTAGACTTTTGAAGAAATATATGATG 649

QY 149

ACGATGCAAAAACATGTTTCAAGAGAGTCTGTAATGCTTTTAAACACACG 208

DB 648

AGATGATGATTTCCAAATTTAGCAGAGAGATGATGAAGATGAAGAAGATGAAG 589

QY 209

TCAACAGAGATTTACTACATGCTGTTGATGTTGAAGATTGACCAAGAAATTTG 268

DB 588

AAGATGAGAGATGATGAAGATGAAGATGAAGATGAAGATGAAGATGAAGATG 529

QY 269

TTGTGACGAAATCTTCACGAAATTTTGTACACAGACTGCGAGGATTAAGATCAAGTA 328

DB 528

AAGAAGATGAAGATGAAGATGAAGATGAAGATGAAGATGAAGATGAAGATGAAG 469

QY 329

ATCGCGCAGTACAGTACAGTACAGTGAAGTTGATGAAGA 371

DB 468

ATGAAGAAGATGAAGATGAAGATGAAGATGAAGATGAAGA 426

RESULT 2

BM168486

LOCUS

701 bp mRNA linear EST 04-DEC-2001

DEFINITION

EST571009 PYBS Plasmodium yoelii yoelii cDNA clone pYCPB36 5' end,

ACCESSION

BM168486

VERSION

BM168486.1

KEYWORDS

GI:17301718

SOURCE

EST.

ORGANISM

Plasmodium yoelii yoelii.

REFERENCE

Plasmodium yoelii yoelii.

AUTHORS

Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

TITLE

Plasmodium yoelii EST project at TIGR

JOURNAL

Unpublished (2001)

COMMENT

Contact: Jane Carlton

Parasite Genomics Group

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-530-9319

Fax: 301-838-0208

FEATURES

Source

Email: carlton@tigr.org
For clone info, please contact the Malaria Research and Reference Reagent Resource Center, ATCC
http://www.malaria.mr4.org/mr4pages/index.html
Seq primer: ADF.

Location/Qualifiers

1..701

/organism="Plasmodium yoelii yoelii"

/strain="17XL"

/db_xref="taxon:73239"

/clone="PYCPB36"

/clone_lib="PYBS"

/dev_stage="Asexual blood stages"

/lab_host="E. coli XL-1 Blue"

/note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was collected from BALB/cByJ mice infected with pY17XL parasites, and leukocytes removed by passage over microcrystalline cellulose columns. Total RNA was isolated using the guanidinium isothiocyanate method, and mRNA isolated using oligo(dT)-cellulose chromatography. First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 5-methyl dCTP. After second strand synthesis, uneven termini were treated with Pfu DNA polymerase and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was precipitated and ligated to HybridZAP arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (PAD-GAL4) was excised from the HybridZAP vector and plasmid DNA isolated."

BASE COUNT

320 a 62 c 147 g 172 t

ORIGIN

Query Match

13.8%; Score 51.6; DB 13; Length 701;

Best Local Similarity

55.6%; Pred. No. 0.00044;

Matches 99; Conservative

0; Mismatches 79; Indels 0; Gaps 0;

QY 198

TGAAGACAGCTCAACAGAGATGATCAATGCTGTTAGTGAAGATTGACCA 257

DB 512

TGAAGATGATGAAGAGATGAAGATGATGACGAAGAGATGAAGATGAAGA 571

QY 258

AGAAAAATTTGTTGTCAGCAAACTTCAAGAAATTTTGTACAGACTGCGAGGTAA 317

DB 572

GGATGAAGATGACGAAGAGATGAAGATGACGAAGATGAAGATGAAGATGA 631

QY 318

AGATGACAGGATGCGGACAGTACAGTACAGTGAAGTGAAGTGAAGTAA 375

DB 632

AGATGAAGATGATGAAGAGATGAAGATGATGAAGATGAAGATGAAGATGA 689

RESULT 3

BM165474

LOCUS

721 bp mRNA linear EST 04-DEC-2001

DEFINITION

EST567997 PYBS Plasmodium yoelii yoelii cDNA clone pYCM68 5' end,

ACCESSION

BM165474

VERSION

BM165474.1

KEYWORDS

GI:17311155

SOURCE

EST.

ORGANISM

Plasmodium yoelii yoelii.

REFERENCE

Plasmodium yoelii yoelii.

AUTHORS

Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

TITLE

Plasmodium yoelii EST project at TIGR

JOURNAL

Unpublished (2001)

COMMENT

Contact: Jane Carlton

Parasite Genomics Group

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-530-9319

Fax: 301-838-0208

Unpublished (2001)
Contact: Jane Carlton

Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208

Email: carlton@tigr.org
For clone info, please contact the Malaria Research and Reference Reagent Resource Center, ATCC
<http://www.malaria.mri4.org/mr4pages/index.html>
led primers and

Seq primer: ADF.	Location/qualifiers
1	

1. 272
/organism="Plasmodium yoelii yoelii"
/strain="17XL"
/db_xref="taxon:73233"
/clone="PYCNC52"
/clone_11b="PyBsn"
/dev_stage="Asexual blood stages"
/lab_host="E. coli XL-1 Blue"
/note="Vector: pAD-XL4; At 20-25% parasitemia, blood was collected from BALB/cByJ mice infected with PY17L parasites, and leukocytes removed by passage over microcrystalline cellulose columns. Total RNA was isolated using the guanidinium isothiocyanate method, and mRNA isolated using oligo(dT)-cellulose chromatography. First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 5-methyl dCTP. After second strand synthesis, uneven seminants were treated with Pfu DNA polymerase and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was precipitated and ligated to HybriZAP arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (pAD-GAL4) was excised from the HybriZAP vector and plasmid DNA

Query Match	13.3%;	Score 50;	DB 13;	Length 272;
Best Local Similarity	55.1%;	Pred. No. 0.0011;		
Matches	08.	Conserved		

[illegible]

BM166886	LOCUS	DEFINITION
BM166886	274 bp	mRNA
EST569409	pybs	Plasmodium yoelli
		yoelli cDNA clone
		pyCZW14 5' end,

ACCESSION BM166886 GI:17300118
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE

EST⁺
Plasmodium yoelii yoelii.
Plasmodium yoelii yoelii.
Eukaryota: Alveolata; Apicomplexa; Haemosporida; Plasmodium.
1 (bases 1 to 274)
Carlton, J.M., Belyut, M., Long, C.A., Bergman, L.W., Valdiva, A.B.
Fischer, C.M., and Carucci, D.J.

JOURNAL
COMMENT

JOURNAL COMMENT
Unpublished (2001)
Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@tigr.org
For clone info, please contact the Malaria Research and Reference
Reagent Resource Center, ARCC
<http://www.malaria.mrt4.org/mr4pages/index.html>
Seq primer: ADP.
Location/Qualifiers

FEATURES

source

1. 276

/organism="Plasmodium yoelii yoelii"

/strain="17XL"

/db_xref="taxon:73239"

/clone="PYCD34"

/clone_lib="PYBS"

/dev_stage="Asexual blood stages"

/lab_host="E. coli XL-1 Blue"

/note="Vector: pAD-GAL4; At 20-25% parasitemia, blood was collected from BALB/CBY1 mice infected with PY17XL parasites, and leukocytes removed by passage over microcrystalline cellulose columns. Total RNA was isolated using the guanidinium isothiocyanate method, and mRNA isolated using oligo(dT)₁₂₋₁₈-cellulose chromatography. First strand cDNA synthesis was completed using a 50 primer and reverse transcriptase in the presence of 5-methyl dCTP. After second strand synthesis, uneven termini were treated with Pfu DNA polymerase and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was precipitated and ligated to HybriZAP arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (pAD-GAL4) was excised from the HybriZAP vector and plasmid DNA isolated."

139 a 12 c 76 g 49 t

BASE COU
ORIGIN

Query Match	13.38	Score 50:	DB 13:	Length 276;
Best Local Similarity	55.18	Pred. No.	0.0011	
Matches	98;	Conservative	0;	Mismatches 80; Indels 0; Gaps 0
QY 198	TCAATACACGGTCAACAAAGAAATGTTACTACTCATGTCGTTGATGGTGAAGAGTTAGACA	257		
Db 93	TGAAGATGATGAGAGAGATGAGATGAGATGATGATGACGAAGAAGATGAAGATGATGATGA	152		
QY 258	AGAAAAATTTGTTGTCGACGAAAACCTTCACGAAAAATTTATTTGACAGACTCGAGGGTAA	317		
Db 153	GGATGAGATGACGAAGAAGATGAGATGACGAGAAAGATGAAGATGACGAAGAAGATGA	212		
QY 318	AGATGACAGTATTCGGCAGCGTACAGGTGACGATCAGATGACGATGATGAAGATTTAA	375		
Db 213	AGATGAGATGATGAGAAGATGAAAGATGATGAAAGAGATGAAGAAGATGAAGATGAA	270		

RESULT

BM162507	425 bp	mRNA	linear	EST 04-DEC-200
LOCUS	BM162507			
DEFINITION	EST65030	pyBS	Plasmodium yoelli	yoelli cDNA clone PYCKR48 5' end,
ACCESSION	mRNA sequence.			
	BM162507			

VERSION
KEYWORD

SOURCE plasmodium yoellii yoellii.
ORGANISM plasmodium yoellii yoellii
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium

REFEREN
AUTHO

Fraser, C.M. and Carucci, D.J.
Plasmodium yoelli EST project at TIGR

JOURNAL

JOURNAL
COMMENT
Unpublished (2001)
Contact: Jane Carlton

Unpublished (2001)
Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Tel: 301-530-9319
Fax: 301-838-0208

RESULTS

http://www.malaria.mr4.org/mr4pages/index.html
Seq primer: ADF.
Location/Qualifiers
1..453

1.453

BASE COUNT

223 a 25 c 112 d 93 +

BASE COUNT	ORIGIN
223 a	25 c
112 g	93 t

Query Match	13 39	Score 50	DP 13	Time 11.1 sec
1	1	100	1	11.1
2	2	100	2	11.1
3	3	100	3	11.1
4	4	100	4	11.1
5	5	100	5	11.1
6	6	100	6	11.1
7	7	100	7	11.1
8	8	100	8	11.1
9	9	100	9	11.1
10	10	100	10	11.1
11	11	100	11	11.1
12	12	100	12	11.1
13	13	100	13	11.1
14	14	100	14	11.1
15	15	100	15	11.1
16	16	100	16	11.1
17	17	100	17	11.1
18	18	100	18	11.1
19	19	100	19	11.1
20	20	100	20	11.1
21	21	100	21	11.1
22	22	100	22	11.1
23	23	100	23	11.1
24	24	100	24	11.1
25	25	100	25	11.1
26	26	100	26	11.1
27	27	100	27	11.1
28	28	100	28	11.1
29	29	100	29	11.1
30	30	100	30	11.1
31	31	100	31	11.1
32	32	100	32	11.1
33	33	100	33	11.1
34	34	100	34	11.1
35	35	100	35	11.1
36	36	100	36	11.1
37	37	100	37	11.1
38	38	100	38	11.1
39	39	100	39	11.1
40	40	100	40	11.1
41	41	100	41	11.1
42	42	100	42	11.1
43	43	100	43	11.1
44	44	100	44	11.1
45	45	100	45	11.1
46	46	100	46	11.1
47	47	100	47	11.1
48	48	100	48	11.1
49	49	100	49	11.1
50	50	100	50	11.1
51	51	100	51	11.1
52	52	100	52	11.1
53	53	100	53	11.1
54	54	100	54	11.1
55	55	100	55	11.1
56	56	100	56	11.1
57	57	100	57	11.1
58	58	100	58	11.1
59	59	100	59	11.1
60	60	100	60	11.1
61	61	100	61	11.1
62	62	100	62	11.1
63	63	100	63	11.1
64	64	100	64	11.1
65	65	100	65	11.1
66	66	100	66	11.1
67	67	100	67	11.1
68	68	100	68	11.1
69	69	100	69	11.1
70	70	100	70	11.1
71	71	100	71	11.1
72	72	100	72	11.1
73	73	100	73	11.1
74	74	100	74	11.1
75	75	100	75	11.1
76	76	100	76	11.1
77	77	100	77	11.1
78	78	100	78	11.1
79	79	100	79	11.1
80	80	100	80	11.1
81	81	100	81	11.1

Best Local Similarity	55.18;	score 30;	DB 13;	Length 453;
Matches 08; Consensus 45;		Pred. No. 0.0012;		

[illegible][illegible]

QY 318 AGATCAGCGTAATGCGGCAGCTACAGGTGACGAGTCAGATGAAGTTGATGAAGATTAA 375
||||| | |||| | |||| |
324

```

DB      2/4  AGATGAAGATCATGAAGAAGATGAGATGATGATGAGAGAGATGAGAAGATGAGATGAA 331
RESULT 11

```

BM171330	LOCUS	DEFINITION
BM171330	459 bp	mRNA
ES7573853	PyBS Plasmodium	linear
	yeoel1	cdna clone
	yeoel1	PCOT11 5' end.
		EST 04-DEC-2001

ACCESSION	miRNA sequence, BM171330
VERSION	BM171330.1
KEYWORDS	GI:17304562 EST.

SOURCE ORGANISM	Plasmodium yoelii Plasmodium yoelii Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
1/1999	1/1999

AUTHORS	TITLE
1 (cases 1 to 459)	
Carlton, J.M., Daly, T.M., Long, C.A., Bergman, L.W., Valdivia, A.B., Fraser, C.M. and Carucci, D.J.	Plasmidium yoeili EST protect at TIGR

JOURNAL COMMENT

Unpublished (2001)
Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@tigr.org
For clone info, please contact the Malaria Research and Reference Reagent Resource Center, ATCC
<http://www.malaria.mr4.org/mr4pages/index.html>
Seq primer: ADF.

FEATURES

source

Location/Qualifiers

1..459

/organism="Plasmodium yoelii yoelii"

/strain="17XL"

/db_xref="taxon:73239"

/clone="PYCQ11"

/clone_lib="PyBS"

/dev_stage="Asexual blood stages"

/lab_host="E. coli XL-1 Blue"

/note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was collected from BALB/cByJ mice infected with Py17XL parasites, and leukocytes removed by passage over microcrystalline cellulose columns. Total RNA was isolated using the guanidium isothiocyanate method, and mRNA isolated using oligo(dT)-cellulose chromatography. First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 5-methyl dCTP. After second strand synthesis, uneven termini were treated with Pfu DNA polymerase and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was precipitated and ligated to HyriZAP arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (PAD-GAL4) was excised from the HyriZAP vector and plasmid DNA isolated."

BASE COUNT 225 a 37 c 113 g 84 t

ORIGIN

Query Match 13.3%; Score 50; DB 13; Length 459;

Best Local Similarity 55.1%; Pred. No. 0.0012;

Matches 98; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 198 TGAAGACAGGTCACAAAGATGTTACTACATGCTGTTGCTGAAGATTAGCCA 257

DB 108 TGAAGATGATGAAGAAGATGAAGATGATGACGAAGAAGATGATGAAGA 167

QY 258 AGAAATTTGTTGTCGACGAAACTTCACGAAATTTTTCACAGACTGCGAGGTAA 317

DB 168 GGATGAAGATGACGAAGAAGATGAAGATGACGAAGAAGATGAAGAAGATGA 227

QY 318 AGATGACAGTAATGCGCAGGTACAGTGCAGTGCAGTGCAGTGCAGTGA 375

DB 228 AGATGAAGATGATGAAGAAGATGAAGATGAAGATGAAGATGAAGATGA 285

RESULT 12

BM161175

LOCUS

DEFINITION 580 bp mRNA linear EST 04-DEC-2001

EST563698 PYBS Plasmodium yoelii yoelii cDNA clone PYCRB54 5' end,

mRNA sequence.

ACCESSION BM161175.1 GI:17306856

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

Plasmodium yoelii yoelii.
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
1 (bases 1 to 580)
Carlton,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Valdivia,A.B., Fraser,C.M. and Carucci,D.J.
Plasmodium yoelii EST project at TIGR

JOURNAL COMMENT

Unpublished (2001)
Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@tigr.org
For clone info, please contact the Malaria Research and Reference Reagent Resource Center, ATCC
<http://www.malaria.mr4.org/mr4pages/index.html>
Seq primer: ADF.

FEATURES

source

Location/Qualifiers

1..580

/organism="Plasmodium yoelii yoelii"

/strain="17XL"

/db_xref="taxon:73239"

/clone="PYCRB54"

/clone_lib="PyBS"

/dev_stage="Asexual blood stages"

/lab_host="E. coli XL-1 Blue"

/note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was collected from BALB/cByJ mice infected with Py17XL parasites, and leukocytes removed by passage over microcrystalline cellulose columns. Total RNA was isolated using the guanidium isothiocyanate method, and mRNA isolated using oligo(dT)-cellulose chromatography. First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 5-methyl dCTP. After second strand synthesis, uneven termini were treated with Pfu DNA polymerase and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was precipitated and ligated to HyriZAP arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (PAD-GAL4) was excised from the HyriZAP vector and plasmid DNA isolated."

BASE COUNT 251 a 47 c 139 g 143 t

ORIGIN

Query Match 13.3%; Score 50; DB 13; Length 580;

Best Local Similarity 55.1%; Pred. No. 0.0012;

Matches 98; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 198 TGAAGACAGGTCACAAAGATGTTACTACATGCTGTTGCTGAAGATTAGCCA 257

DB 397 TGAAGATGATGAAGAAGATGAAGATGATGACGAAGAAGATGATGAAGA 456

QY 258 AGAAATTTGTTGTCGACGAAACTTCACGAAATTTTTCACAGACTGCGAGGTAA 317

DB 457 GGATGAAGATGACGAAGAAGATGAAGATGACGAAGAAGATGAAGAAGATGA 516

QY 318 AGATGACAGTAATGCGCAGGTACAGTGCAGTGCAGTGCAGTGCAGTGA 375

DB 517 AGATGAAGATGATGAAGAAGATGAAGATGAAGATGAAGATGAAGATGA 574

RESULT 13

BM164042

LOCUS

DEFINITION 648 bp mRNA linear EST 04-DEC-2001

EST56565 PYBS Plasmodium yoelii yoelii cDNA clone PYC1P02 5' end,

mRNA sequence.

ACCESSION BM164042.1 GI:17309723

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

Plasmodium yoelii yoelii.
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
1 (bases 1 to 648)
Carlton,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Valdivia,A.B., Fraser,C.M. and Carucci,D.J.
Plasmodium yoelii EST project at TIGR

Wed Feb 26 15:58:58 2003

us-09-936-737a-1.rst

JOURNAL
COMMENT

Unpublished (2001)
Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-550-5119
Fax: 301-838-0208
Email: jcarlton@broadinstitute.org

Email: carlton@tigr.org
For clone info, please

Reagent Resource Center, ATCC
http://www.malaria.mr4.org/mr4pages/index.html
Seq primer: ADF.

FEATURES

Source

Location/Qualifiers
1. .667

```
/originals="Plasmodium yoelii yoelii"  
/strain="17XL"
```

```
/add_xref="taxon:73239"
/clone="pvcvrd00"
```

```
/clone_lib="pyBS"
```

```
/dev_stage="Asexual blood stages"
```

```

... coll AL-1 Blue"
/note="Vector: PAD-GAL4: At 2

```

collected from BALB/cBJ mice infected with py17XL parasites, and leukocytes removed by passage

microcrystalline cellulose columns. Total RNA was isolated using the guanidinium isothiocyanate method.

First strand cDNA synthesis was completed using a first strand cDNA synthesis kit (Stratagene, La Jolla, CA) using oligo(dT)-cellulose chromatography, and reverse transcriptase (RT) (Stratagene, La Jolla, CA) using oligo(dT)-cellulose chromatography, and reverse transcriptase (RT) (Stratagene, La Jolla, CA).

5-methyl dCTP. After second strand synthesis in the presence of primer and reverse transcriptase using a 50-base

adaptors ligated to the blunt ends. The sample was cleaved

Size-fractionated cDNA was precipitated and ligated to hybridization arms attached to Sepharose 4B beads and separated on a Sepharose 4B column.

After packaging, the phagemid vector (PAD-GAL4) was excised from the λ -phage and cleaved arms.

the hybridization vector and plasmid DNA isolated.

BASE COUNT	301 a	58 c	145 g	163 t
ORIGIN				

Query Match

Exact match	13.3%	Score 50;	DB 13;	Length 667;
Best Local Similarity	55.18%	Pred No	0	0013

malenes 9

	Indels	Gaps
198 TGAGGACACCCTCCGGCGGC.....	0;	0;

484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527
 528
 529
 530
 531
 532
 533
 534
 535
 536
 537
 538
 539
 540
 541
 542
 543
 544
 545
 546
 547
 548
 549
 550
 551
 552
 553
 554
 555
 556
 557
 558
 559
 560
 561
 562
 563
 564
 565
 566
 567
 568
 569
 570
 571
 572
 573
 574
 575
 576
 577
 578
 579
 580
 581
 582
 583
 584
 585
 586
 587
 588
 589
 590
 591
 592
 593
 594
 595
 596
 597
 598
 599
 600
 601
 602
 603
 604
 605
 606
 607
 608
 609
 610
 611
 612
 613
 614
 615
 616
 617
 618
 619
 620
 621
 622
 623
 624
 625
 626
 627
 628
 629
 630
 631
 632
 633
 634
 635
 636
 637
 638
 639
 640
 641
 642
 643
 644
 645
 646
 647
 648
 649
 650
 651
 652
 653
 654
 655
 656
 657
 658
 659
 660
 661
 662
 663
 664
 665
 666
 667
 668
 669
 670
 671
 672
 673
 674
 675
 676
 677
 678
 679
 680
 681
 682
 683
 684
 685
 686
 687
 688
 689
 690
 691
 692
 693
 694
 695
 696
 697
 698
 699
 700
 701
 702
 703
 704
 705
 706
 707
 708
 709
 710
 711
 712
 713
 714
 715
 716
 717
 718
 719
 720
 721
 722
 723
 724
 725
 726
 727
 728
 729
 730
 731
 732
 733
 734
 735
 736
 737
 738
 739
 740
 741
 742
 743
 744
 745
 746
 747
 748
 749
 750
 751
 752
 753
 754
 755
 756
 757
 758
 759
 760
 761
 762
 763
 764
 765
 766
 767
 768
 769
 770
 771
 772
 773
 774
 775
 776
 777
 778
 779
 780
 781
 782
 783
 784
 785
 786
 787
 788
 789
 790
 791
 792
 793
 794
 795
 796
 797
 798
 799
 800
 801
 802
 803
 804
 805
 806
 807
 808
 809
 810
 811
 812
 813
 814
 815
 816
 817
 818
 819
 820
 821
 822
 823
 824
 825
 826
 827
 828
 829
 830
 831
 832
 833
 834
 835
 836
 837
 838
 839
 840
 841
 842
 843
 844
 845
 846
 847
 848
 849
 850
 851
 852
 853
 854
 855
 856
 857
 858
 859
 860
 861
 862
 863
 864
 865
 866
 867
 868
 869
 870
 871
 872
 873
 874
 875
 876
 877
 878
 879
 880
 881
 882
 883
 884
 885
 886
 887
 888
 889
 890
 891
 892
 893
 894
 895
 896
 897
 898
 899
 900
 901
 902
 903
 904
 905
 906
 907
 908
 909
 910
 911
 912
 913
 914
 915
 916
 917
 918
 919
 920
 921
 922
 923
 924
 925
 926
 927
 928
 929
 930
 931
 932
 933
 934
 935
 936
 937
 938
 939
 940
 941
 942
 943
 944
 945
 946
 947
 948
 949
 950
 951
 952
 953
 954
 955
 956
 957
 958
 959
 960
 961
 962
 963
 964
 965
 966
 967
 968
 969
 970
 971
 972
 973
 974
 975
 976
 977
 978
 979
 980
 981
 982
 983
 984
 985
 986
 987
 988
 989
 990
 991
 992
 993
 994
 995

484 TGAGATGATGAGAGATGAGATGAGATGATGAGAGAGAGATGAGATGATGAAGA 543

258 AGAAAAATTTGTTGTCGACGAAACTTCACGGAAATATTGACACACCTCCCGCGCGCTT

544 GGATGAGATGACGAAGAAGATGAGATCACTCCCGGGTAA
||||||| |
317

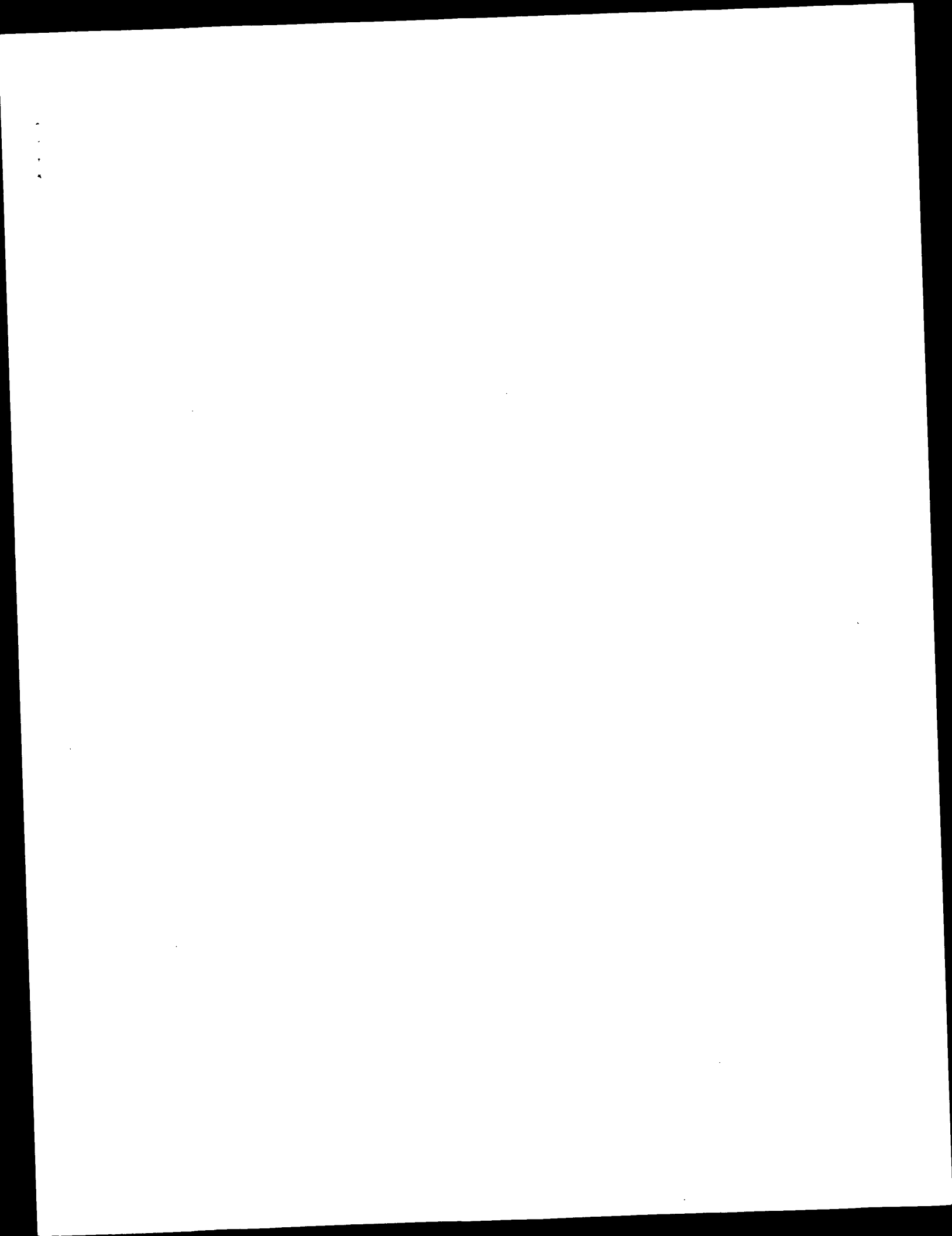
318 AGATGCAGGTAATCCCCCAACCTAATT
603

||||| |||| | ||| |||||
604 AGATGAGTGGACGGTACAGGTGACGAGTCAGATGAAGTTGATGACAGATTAA 375
ACATGAGTGGACGGTACAGGTGACGAGTCAGATGAAGTTGATGACAGATTAA

504 AGAIGAAATGATGAAGAGATGAAGATGATGAAGAGGATGAAGAGAGATGAGATGAA 661

Search completed: February 25, 2003, 02:23:06
Job time : 1967 secs

Page 9



Run on:

February 25, 2003, 02:28:20 ; Search time 27 Seconds

(without alignments)
508.326 Million cell updates/sec

Title: US-09-936-737A-2
Perfect score: 573
Sequence: 1 EEREDCWTFYANKRYTDFDK

Scoring table: BLOSUM62

gapop 10.0 , gapext 0.5

133250620 residues

Verifying chosen parameters: 908470

```
Maximum DB seq length: 0
Maximum DB seq length: 2000000000
```

```
post-processing: Minimum match 0%
```

Maximum Match 100%

listing first 45 summaries

Database :

1: *SIDS27.gcgdata.geneseq.geneseqp.embl.AA1980.DAT.**
2: *SIDS27.gcgdata.geneseq.geneseqp.embl.AA1980.DAT.**
3: *SIDS27.gcgdata.geneseq.geneseqp.embl.AA1981.DAT.**
4: *SIDS27.gcgdata.geneseq.geneseqp.embl.AA1981.DAT.**
5: *SIDS27.gcgdata.geneseq.geneseqp.embl.AA1983.DAT.**
6: *SIDS27.gcgdata.geneseq.geneseqp.embl.AA1984.DAT.**
7: *SIDS27.gcgdata.geneseq.geneseqp.embl.AA1985.DAT.**
8: *SIDS27.gcgdata.geneseq.geneseqp.embl.AA1986.DAT.**
9: *SIDS27.gcgdata.geneseq.geneseqp.embl.AA1987.DAT.**
10: *SIDS27.gcgdata.geneseq.geneseqp.embl.AA1988.DAT.**
11: *SIDS27.gcgdata.geneseq.geneseqp.embl.AA1989.DAT.**
12: *SIDS27.gcgdata.geneseq.geneseqp.embl.AA1990.DAT.**
13: *SIDS27.gcgdata.geneseq.geneseqp.embl.AA1991.DAT.**
14: *SIDS27.gcgdata.geneseq.geneseqp.embl.AA1992.DAT.**
15: *SIDS27.gcgdata.geneseq.geneseqp.embl.AA1993.DAT.**
16: *SIDS27.gcgdata.geneseq.geneseqp.embl.AA1994.DAT.**
17: *SIDS27.gcgdata.geneseq.geneseqp.embl.AA1995.DAT.**
18: *SIDS27.gcgdata.geneseq.geneseqp.embl.AA1996.DAT.**
19: *SIDS27.gcgdata.geneseq.geneseqp.embl.AA1998.DAT.**
20: *SIDS27.gcgdata.geneseq.geneseqp.embl.AA2000.DAT.**
21: *SIDS27.gcgdata.geneseq.geneseqp.embl.AA2000.DAT.**
22: *SIDS27.gcgdata.geneseq.geneseqp.embl.AA2001.DAT.**
23: *SIDS27.gcgdata.geneseq.geneseqp.embl.AA2002.DAT.**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query Match	Length	DB	ID	Description
	1	573	100.0	103	21	AA94746	Platelet binding
	2	80.3	14.0	147	13	AAR2250	Leech antiplatelet
	3	80.5	14.0	412	20	AAM68010	S. frugiperda immu
	4	77	13.4	368	22	AAB59867	Drosophila melano
	5	75.5	13.2	418	21	AAB1260	Mouse Ics-4 prote
	6	75.5	13.2	419	20	AAB50109	Murine caspase-12
	7	75.5	13.2	419	21	AAB14259	Mouse Ics-4 prote
	8	75.5	13.2	419	21	AAB29108	Murine caspase-12
	9	75.5	13.2	419	21	AAB29108	Murine caspase-12
	10	75.5	13.2	432	16	AAR66769	Human Interleukin
						AA98464	Murine Ics-3. Mus

11	73.5	12.8	1154	22	ABB66916
12	71.5	12.5	175	21	AAG17789
13	71.5	12.5	256	21	AAG17788
14	71.5	12.5	320	21	AAB18963
15	71	12.4	589	22	AAB95086
16	70.5	12.3	156	23	AA699411
17	70.5	12.3	367	22	ABB71954
18	70	12.2	669	23	ABB09771
19	70	12.2	669	23	ABB09781
20	70	12.2	1167	22	AAU35929
21	69.5	12.1	674	19	AAAP54656
22	69.5	12.1	674	23	AAAP54656
23	69.5	12.1	3257	22	ABB67502
24	69	12.0	717	21	ABAB1819
25	69	12.0	717	22	AAAB3298
26	69	12.0	734	22	ABB11943
27	69	12.0	734	22	AAAB91943
28	69	12.0	734	22	AAAB93299
29	69	12.0	734	22	AAAB94096
30	69	12.0	738	22	AAAM1084
31	68.5	12.0	738	22	AAAM1085
32	68.5	12.0	610	23	ABAG61895
33	68.5	12.0	635	21	AAAB18176
34	68	11.9	1869	22	ABBS8651
35	68	11.9	146	23	ABBO8153
36	68	11.9	403	22	AAAB94657
37	68	11.9	561	20	AAAY41665
38	68	11.9	867	22	ABG07270
39	68	11.9	903	22	AAAB39516
40	67.5	11.8	961	22	AAAB78435
41	67.5	11.8	117	23	ABBP08662
42	67.5	11.8	366	21	AAAG53207
43	67.5	11.8	450	21	AAAG34636
44	67.5	11.8	450	21	AAAG53206
45	67.5	11.8	488	21	AAAG34635
			488	21	AAAG53205
					Arabidopsis thaliana
					Drosophila melanogaster
					Arabidopsis thaliana
					Arabidopsis thaliana
					Plasmodium falciparum
					Human protein sequence
					Lung small cell carcinoma
					Drosophila melanogaster
					Amino acid sequence
					Amino acid sequence
					Helicobacter pylori
					Streptococcus pneumoniae
					S. pneumoniae Spili
					Drosophila melanogaster
					Human ORFX ORF1583
					Human polypeptide
					Human d4475B7.2 ho
					Human polypeptide
					Human protein sequence
					Human polypeptide
					Human polypeptide
					Prostate cancer-associated
					Plasmodium falciparum
					Drosophila melanogaster
					Murine PLA2 enzyme
					Human protein sequence
					Zea mays AWC-11 protein
					Novel human diaphanous
					Human polypeptide
					Human protein sequence
					Human ORFX protein
					Arabidopsis thaliana
					Arabidopsis thaliana
					Arabidopsis thaliana
					Arabidopsis thaliana

ALIGNMENTS

RESULT 1
AAIY94746
ID AAIY94746 standard; Protein; 103 AA.
XX

22 JAN-2001 (first entry)
XX

Saratin: modified amino acid sequence

	collagen-dependent platelet adhesion inhibition; intraocular lens;
KW	posterior capsule opacification.
XX	

03 Alrudo medicinalis.
XX

WO200056885-A1.

PD 28-SEP-2000.
YY

10-MAR-2000; 2000WO-EP02117.

18-MAR-1999; 99EP-0105530.

[illegible]

... ..

Scheuble B, Hufmann U, Hemberger J, Fotev Z, Guessow D, Hofmann U

WPI; 2000-611629/58

N-PSDB; AAA28180.

PT New Saratin polypeptide and gene isolated from Hirudo medicinalis for
 PT blocking platelet adhesion, especially useful for treating or
 PT preventing thrombotic diseases, or for manufacturing a medicament for
 PT thromboembolic diseases

PS Claim 7: Page 41; 46pp; English.

CC Saratin is a protein isolated from the saliva of the medicinal leech
 CC Hirudo medicinalis. Saratin is an inhibitor of collagen-dependent
 CC platelet adhesion. The invention includes polynucleotide sequences
 CC encoding Saratin, an expression vector comprising the DNA sequence, a
 CC host cell transformed with the expression vector, antibodies
 CC immunospecific for Saratin, and methods for identifying Saratin agonists
 CC or antagonists. Saratin is useful for treating thromboembolic diseases,
 CC and for manufacturing a medicament for treating thromboembolic diseases.
 CC It is useful for preventing thrombotic diseases. Saratin is also useful
 CC for coating artificial surfaces, since use of Saratin renders them
 CC non-adhesive for cells and prevents the activation of cells. It may also
 CC be used for coating natural collagen surfaces. Furthermore, Saratin is
 CC useful for modifying intraocular lenses in order to lessen the
 CC thrombogenicity of the lens material, for contacting the lens surface, or
 CC for covalent crosslinking to modify the lens material. The lens material
 CC is used for refractive anterior or posterior chamber ocular implants,
 CC which may be implanted into the eye. This new type of coating avoids
 CC problems contributed by stimulated cell growth. In combination with other
 CC medicaments that are for instance conferring cell death, Saratin coating
 CC helps to completely overcome posterior capsule opacification. The
 CC antibody immunospecific for Saratin, as well as Saratin itself, are
 CC useful for measuring samples derived from host cell cultures or from a
 CC treated subject. The present sequence represents the amino acid sequence
 CC of the Saratin protein.

SQ Sequence 103 AA:

Query Match 100.0%; Score 573; DB 21; Length 103;
 Best Local Similarity 100.0%; Pred. No. 2,3e-52; Indels 0; Gaps 0;
 Matches 103; Conservative 0; Mismatches 0;

QY 1 EREDCTFTYANKRYTDFKSFKSSDLDECKKTCFTEYCYIVEDYVNEKCYINVDG 60
 DB 1 EREDCTFTYANKRYTDFKSFKSSDLDECKKTCFTEYCYIVEDYVNEKCYINVDG 60
 QY 61 EELDOEFVNDENTFTENTLTDCEGKDGAGNAAGTDESDEVED 103
 DB 61 EELDOEFVNDENTFTENTLTDCEGKDGAGNAAGTDESDEVED 103

RESULT 2

AAR22950
 ID AAR22950 standard; Protein: 147 AA.

XX AAR22950;

XX 10-SEP-1992 (first entry)

XX Leech antiplatelet protein.

XX LAPP; collagen; platelet aggregation; antithrombotic; anticoagulant;

KW coronary artery disease; cerebrovascular disease; ss.

XX Haementeria officinalis.

XX Location/Qualifiers

XX Key 1..21

XX Peptide /label= leader peptide amino acids"

XX /note= "21 hydrophobic

XX Protein 22..147

XX /label= mature protein

XX EP480651-A.

XX 15-APR-1992.

PF 07-OCT-1991; 91EP-0309157.

XX 09-OCT-1990; 90US-0594917.

XX (MERI) MERCK & CO INC.

XX Connolly TN, Keller PM;

XX WPI: 1992-125288/16.

DR N-PSDB; AAQ23682.

PT New protein to inhibit collagen stimulated platelet aggregation -
 PT extracted from salivary gland of Haementeria officinalis for
 PT treating thrombotic diseases

PS Claim 6: Page 15; 20pp; English.

CC When cloned into the yeast expression vector pKH42, the LAPP gene
 CC sequence forms a fusion product of MA α -alpha-1 leader sequence
 CC plus mature LAPP, which when produced are proteolytically processed by
 CC the Lys-Arg cleaving endopeptidase (KEX2), which cleaves on the
 CC C terminal side of Lys-Arg, and the products secreted into the culture
 CC medium.
 CC LAPP blocks stimulation of platelet aggregation by collagen, at an
 CC antithrombotic concentration of about 100 nMolar (or 1.6 ug/ml).
 CC platelet aggregation is stimulated by 2 ug/ml collagen and addition
 CC of LAPP inhibits this with an IC50 of 45nM, but this may be
 CC overcome by the addition of 0.25 mM arachidonic acid.
 CC LAPP could be used for treatment and prevention of thrombotic
 CC conditions especially coronary artery and cerebrovascular disease
 CC in mammals.
 CC See also AAQ23871-7.

SQ Sequence 147 AA:

Query Match 14.0%; Score 80.5; DB 13; Length 147;
 Best Local Similarity 32.7%; Pred. No. 1,1; Indels 3; Gaps 1;
 Matches 18; Conservative 10; Mismatches 24;

QY 5 DCWTYANKRYTDFKSFKSSDLDECKKTCFTEY---CYIVEDYVNEKCYIN 56
 DB 66 DCWSRPGWKLPDNLITFTSTYVDECKKNCESAVPSCYIIQINTNECYRN 120

RESULT 3

AAW68010
 ID AAW68010 standard; Protein: 412 AA.

XX AAW68010;

XX 27-APR-1999 (first entry)

XX S. frugiperda immunophilin FKBP46.

XX Immunophilin; moth; insect cell; nuclear; immunosuppression; drug;

KW transplant; tissue graft.

XX Spodoptera frugiperda.

XX Key 111..114

XX Domain /note= "EEAP motif"

XX Region 124

XX /note= "putative nuclear localisation signal"

XX Region 135..138

XX /note= "putative nuclear localisation signal"

XX Domain 148..149

XX /note= "putative nuclear localisation signal"

XX /note= "AP motif"

XX Domain 215..218

XX /note= "EEAP motif"

XX /note= "putative nuclear localisation signal"

FT	Region	250..256	/note="putative nuclear localisation signal"
FT	Region	272..275	/note="putative nuclear localisation signal"
FT	Region	284..286	/note="putative nuclear localisation signal"
FT	Domain	292..295	/note="EAP motif"
FT	Region	298..302	/note="putative nuclear localisation signal"
XX			
PN	US5861498-A.		
PD	19-JAN-1999.		
XX			
PF	31-OCT-1996;	96GS-0741134.	
XX			
PR	01-NOV-1995;	95US-0007163.	
PR	31-OCT-1996;	96US-0741134.	
XX	(UYJE-) UNITV JEFFERSON THOMAS.		
FA	Alnemri ES, Fernandes-Alnemri T, Litwack G;		
PI	WPI; 1999-130433/11.		
DR	N-PSTD; AAX04441.		
XX	Isolated immunophilin FKBP46 nucleic acids - useful for developing products for the study and identification of immunosuppressive agents for treating e.g. transplant and tissue graft patients		
PS	Claim 1: Fig 1: 21pp; English.		
CC	This sequence represents the immunophilin FKBP46 from the moth Spodoptera frugiperda. FKBP46 is a novel insect cell nuclear immunosuppressive drug that can be used to study and identify additional immunosuppressive drugs that bind to it. Such immunosuppressive agents can be used in treating transplant and tissue graft patients.		
SQ	Sequence	412 AA:	
OY	Query Match	14.0%: Score 80.5; DB 20; Length 412;	
DB	Best Local Similarity	30.6%; Pred. No.3.9;	
Matches	30; Conservative	12; Mismatches	29; Indels
Gaps			27; Gaps
OY	11 ANRKYTDPEKSEFKS-----SDIDECKTCFETFCYIYFEDTVNRKECYVNVVDGEELDQ 65 : : : 133 ANKKAKPKKKGKNSAPAESDSDDDD-----EQQLK-----FLDGEDI DT 174		
OY	66 EKFVYDENFTENYLTLDEGGKAGNAAGTGDSESDVEDD 103 :: : 175 DE--NDESEKKN--TSAGGDSDEDEDDEDEDED DD 208		
RESULT 4	ABBE9867		
ID	ABBE9867 standard; Protein: 368 AA.		
AC	ABBE9867;		
DT	26-MAR-2002 (first entry)		
DE	Drosophila melanogaster polypeptide SEQ ID NO 36393.		
KM	Drosophila; developmental biology; cell signalling; insecticide;		
OS	pharmaceutical.		
PN	Drosophila melanogaster.		
PD	WO200171042-A2.		
PD	27-SEP-2001.		
XX			

PF 23-MAR-2001; 2001WO-US09231.
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
PA
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX
XX WPI; 2001-656860/75.
DR N-PSDB; ABL13970.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Disclosure; SEQ ID NO 36393; 21pp + Sequence Listing; English.
PS
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL01840-ABL16175) and the encoded proteins
CC sequences (AB057737-AB052072).
CC (AB057737-AB052072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct-sequences.
XX
XX Sequence 368 AA;
XX

[illegible]

DR WPI: 2000-464343/40.

XX New human Ich-1L and Ich-1S proteins for negative and positive
PT regulation of programmed cell death and for developing therapeutic
PT methods for diseases and conditions characterized by cell death, e.g.
XX myocardial infarction or stroke
PS Disclosure; Fig 17; 121pp; English.

XX The present sequence is a mouse Ice-4 protein sequence. The present
CC sequence was used in a sequence homology comparison with the present
CC Interleukin-1beta converting enzyme (ICE) (AAB14253), murine
CC (AAB14253) and C. elegans ced-3 (AAB14246). The coding sequence of the
CC present sequence is a member of a family of genes involved in programmed
CC cell death (apoptosis). Ich-1 may play an important role in both the
CC positive and negative regulation of apoptosis. The Ich gene may be used
CC in gene therapy in disorders characterized by cell death e.g. neural and
CC muscular degenerative diseases, myocardial infarction, stroke, vitally
CC induced cell death and aging.

SO Sequence 418 AA;

Query Match

Best Local Similarity 13.2%; Score 75.5; DB 21; Length 418;
Matches 26; Conservative 14; Mismatches 14; Indels 19; Gaps 6;

OY 44 VFEDTVNKECYNVVDGEEL---DQKRVVD--ENFTENYL--TDCEGR-DAGNAGTG 94

DB 26 VFDDLVK---NVLNDELTKIGESASPIIKAKALNVLNFTKTDMAKIFAGHINSQ 81

OY 95 DE-----SDEVD 101

DB 82 EQLSLQPSNDEDD 94

RESULT 6

AA550109

ID AAY50109 standard; Protein; 419 AA.

AC AAY50109;

DT 21-JAN-2000 (first entry)

DE Murine caspase-12 splice variant, caspase-12L.

XX Caspase; splice variant; truncated; programmed cell death; apoptosis;
KW regulation; proteolytic cascade; malignant condition; cancer;
KW premalignant condition; solid tumor; lymphoma;
KW chronic lymphocytic leukaemia; prostatic hypertrophy;
KW preneoplastic liver focus; chemotherapy resistance; autoimmune disease;
XX ds.

OS Mus sp.

XX Key Location/Qualifiers

FT Cleavage-site 94..95

FT Protein 95..419

FT Cleavage-site 318..319

PN W09952925-A1.

PD 21-OCT-1999.

PF 14-APR-1999; 99WO-US08064.

PR 16-APR-1998; 98US-0081962.

PA (GEHO) GEN HOSPITAL CORP.

PI Yuan J, Morishima N.

DR WPI: 1999-620369/53.

DR N-PSDB: AAZ32645.

XX New nucleic acid encoding the short form of caspase-12, used e.g. for
PT treating tumors
XX

PS Claim 1; Fig 1; 68pp; English.

XX This sequence represents a cDNA encoding murine caspase-12L, a splice
CC variant of caspase-12. Caspases are a family of proteins involved in
CC the regulation of apoptosis and are synthesized as proforms which are
CC activated via cleavage after specific Asp residues. Mammalian cells
CC express several caspases, and it is thought that these act in a
CC proteolytic cascade to cause programmed cell death. Nucleic acids
CC encoding caspase-12S (AA32644) or truncated forms of caspase-12L
CC (AA32646, AA32647) are used for production, recombinantly or in vivo,
CC of caspase-12 polypeptides which induce programmed cell death. This is
CC particularly useful for treating (pre)malignant conditions (e.g. solid
CC tumours, B cell lymphoma, chronic lymphocytic leukaemia, prostatic
CC hyperplasia, preneoplastic liver foci and resistance to chemotherapy), or
CC autoimmune diseases. The caspase-12 proteins can also be used to raise
CC specific antibodies (for example, to determine gene expression and to
CC screen expression libraries) or as molecular weight markers. Fragments
CC of caspase-12 gene (and its allelic variants) can be used as probes to isolate
CC hybridisation for chromosomal location of the caspase-12 gene, and for
CC Northern blotting to determine caspase-12 mRNA expression in tissues.

SO Sequence 419 AA;

Query Match

Best Local Similarity 13.2%; Score 75.5; DB 20; Length 419;
Matches 26; Conservative 14; Mismatches 14; Indels 19; Gaps 6;

OY 44 VFEDTVNKECYNVVDGEEL---DQKRVVD--ENFTENYL--TDCEGR-DAGNAGTG 94

DB 26 VFDDLVK---NVLNDELTKIGESASPIIKAKALNVLNFTKTDMAKIFAGHINSQ 81

OY 95 DE-----SDEVD 101

DB 82 EQLSLQPSNDEDD 94

RESULT 7

AAB14259

ID AAB14259 standard; Protein; 419 AA.

AC AAB14259;

DT 09-FEB-2001 (first entry)

DE Mouse Ice-4 protein sequence #1.

XX ced-3; virally induced cell death; apoptosis; gene therapy; neural;
KW muscular degenerative disease; myocardial infarction; stroke; aging;
KW Interleukin-1beta converting enzyme; ICE; cysteine protease; mouse.
XX

OS Mus sp.

PN US6083735-A.

PD 04-JUL-2000.

PF 10-JUN-1994; 94US-0258287.

PR 24-JUN-1993; 93US-0080850.

PA (GEHO) GEN HOSPITAL CORP.

PI Yuan J, Miura M.

DR WPI: 2000-464343/40.

DR N-PSDB: AAA72843.

Query Match	13.2%;	Score 75.5;	DB 21;	Length 419;
Best Local Similarity	35.6%;	Pred. No. 13;		
Matches 26;	Conservative 14;	Mismatches 14;	Indels 19;	Gaps 6;
44	VFEEDYVNEKCYNVVGGEL----	DDKRPVVD--ENFTENTLY--TDDEGR-DAGNAACTG	94	
26	VDDDLVER-----NVLNGDELKIGESAFILINKAENVLENTLEKTDMAKGFIGHIANSO	81		
95	DE-----SDEVD	101		
82	EQLSLQFSDNEDD	94		
<p>RESULT 9</p> <p>AA66769</p> <p>ID AA66769 standard; Protein: 432 AA.</p> <p>AA66769:</p> <p>12-SEP-1995 (first entry)</p> <p>Human interleukin-1 beta converting enzyme homolog (Ice-4).</p> <p>Human interleukin-1 beta converting enzyme homolog; Ice-4;</p> <p>oncogene bcl-2; programmed cell death; cancer treatment.</p> <p>Homo sapiens.</p> <p>W09500160-A.</p> <p>05-JAN-1995.</p> <p>10-JUN-1994; 94WO-US06630.</p> <p>24-JUN-1993; 93US-0080850.</p> <p>(GENE) GEN HOSPITAL CORP.</p> <p>Mura M. Yuan J;</p> <p>WPI: 1995-051742/07.</p> <p>N-PSDB; AA079969.</p> <p>Promoting or preventing programmed cell death in vertebrate cells</p> <p>- by inhibiting the activity of interleukin-1 beta converting enzyme.</p> <p>Claim 24; Fig 16; 116pp; English.</p> <p>AA079969 encodes AA66769 human interleukin-1 beta converting enzyme homolog (Ice-4), increasing Ice-4s enzymatic activity can promote the programmed cell death of cancer cells (pref. those overexpressing the bcl-2 oncogene), this can be used as the basis of a new cancer treatment. Alternatively by reducing Ice-4s enzymatic activity programmed cell death can be inhibited, this may be useful in the development of new cell lines which remain viable in culture for extended or indefinite periods, independant of growth factors.</p> <p>Sequence 432 AA;</p> <p>Query Match 13.2%; Score 75.5; DB 16; Length 432;</p> <p>Best Local Similarity 35.6%; Pred. No. 14;</p> <p>Matches 26; Conservative 14; Mismatches 14; Indels 19; Gaps 6;</p>				

OY 44 VFEDTVNKECYNVVGEEL----DQKRVVD--ENFTENYL--TDCEGK-DAGNAGTG 94
 DB 40 VFEDTVNKECYNVVGEEL----DQKRVVD--ENFTENYL--TDCEGK-DAGNAGTG 94
 OY 95 DE-----SDEVD 101
 DB 96 EQLSIQFSNDEDD 108

RESULT 10
 AAR98464
 ID AAR98464 standard; Protein; 432 AA.
 AC AAR98464;
 XX
 DT 25-SEP-1996 (first entry)
 DE Murine Ich-3.
 DE
 KW Ich-3; murine ICE-ced-3 homologue; programmed cell death;
 KM apoptosis; interleukin-1 beta converting enzyme; gene therapy.
 XX
 OS Mus sp.
 PN WO9620721-A1.
 PD 11-JUL-1996.
 PF 04-JAN-1996; 96MO-US00177.
 PR 04-JAN-1995; 95US-0368704.
 PA (GENO) GEN HOSPITAL CORP.
 PI Miura M, Yuan J;
 DR WPI: 1996-333763/33.
 DR N-PSDB: AAT31554.

Preventing or promoting programmed cell death in vertebrate cells -
 comprises inhibiting or increasing the activity of
 interleukin-1-beta converting enzyme, or altering expression of
 other related genes
 Claim 24: Fig 14; 127pp; English.
 CC Ich-3 (AAR98464) causes programmed cell death and shows significant
 CC homology to mouse interleukin-1 beta converting enzyme (ICE),
 CC mouse mitch-2 (AAR98461) and human Ich-1 (AAR98462-63). Its sequence
 CC was deduced from the Ich-3 gene (AAT31554) isolated from a mouse thymus
 CC cDNA library. The protein can be obtained from host cells containing
 CC vectors that include an Ich-3 coding sequence. It can be used to
 CC control the programmed cell death of vertebrate cells, to develop cell
 CC lines that remain viable for extended periods, and to increase the
 CC activity of ICE.
 XX

Sequence 432 AA;
 Query Match 13.2%; Score 75.5; DB 17; Length 432;
 Best Local Similarity 35.6%; Pred. No. 14;
 Matches 26; Conservative 14; Mismatches 14; Indels 19; Gaps 6;

OY 44 VFEDTVNKECYNVVGEEL----DQKRVVD--ENFTENYL--TDCEGK-DAGNAGTG 94
 DB 40 VFEDTVNKECYNVVGEEL----DQKRVVD--ENFTENYL--TDCEGK-DAGNAGTG 94
 OY 95 DE-----SDEVD 101
 DB 96 EQLSIQFSNDEDD 108

RESULT 11
 ABB66916

ID ABB66916 standard; Protein; 1154 AA.
 XX
 AC ABB66916;
 XX
 DT 26-MAR-2002 (first entry)
 DE Drosophila melanogaster polypeptide SEQ ID NO 27540.
 DE
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KM pharmaceutical.
 XX
 OS Drosophila melanogaster.
 PN WO200171042-A2.
 PD 27-SEP-2001.
 PF 23-MAR-2001; 2001MO-US09231.
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 PA (PEKE) PE CORP NY.
 PI Venter JC, Adams M, Li PWD, Myers EW;
 DR WPI: 2001-656860/75.
 DR N-PSDB: ABL11019.

New isolated nucleic acid detection reagent for detecting 1000 or more
 genes from Drosophila and for elucidating cell signalling and cell-cell
 interactions -
 Disclosure: SEQ ID NO 27540; 21pp + sequence listing; English.
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at http://wipo.int/pub/published_pct_sequences.
 XX

Sequence 1154 AA;
 Query Match 12.8%; Score 73.5; DB 22; Length 1154;
 Best Local Similarity 28.1%; Pred. No. 73;
 Matches 27; Conservative 14; Mismatches 42; Indels 13; Gaps 4;

OY 12 NRRYDFDKSEKSSDLECK-KTCTFTEYCYIVFEDTVNKECYNVVYDGEEDQKRV 69
 DB 505 SKRDIKSEKSSDLECK-KTCTFTEYCYIVFEDTVNKECYNVVYDGEEDQKRV 69
 OY 70 VDNEFTENYL--TDCEGDA-----GNAAGTG 94
 DB 565 IVENFWRHRYICCCQKQKACALSNPSMIGNTERTYTG 600

RESULT 12
 AAG17789
 ID AAG17789 standard; Protein; 175 AA.
 AC AAG17789;
 XX
 DT 17-OCT-2000 (first entry)
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 18943.
 DE
 KW Protein identification; signal transduction pathway; metabolic pathway;

hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence.
Arabidopsis thaliana.
EP1033405-A2.
06-SEP-2000.
25-FEB-2000; 2000EP-0301439.
25-FEB-1999; 99US-0121825.
05-MAR-1999; 99US-0123180.
09-MAR-1999; 99US-0123548.
23-MAR-1999; 99US-0125788.
25-MAR-1999; 99US-0126264.
29-MAR-1999; 99US-0126785.
01-APR-1999; 99US-0127462.
06-APR-1999; 99US-0128234.
08-APR-1999; 99US-0128714.
16-APR-1999; 99US-0129845.
19-APR-1999; 99US-0130077.
21-APR-1999; 99US-0130449.
23-APR-1999; 99US-0130510.
28-APR-1999; 99US-0130891.
30-APR-1999; 99US-0131449.
30-APR-1999; 99US-0132048.
04-MAY-1999; 99US-0132407.
05-MAY-1999; 99US-0132484.
06-MAY-1999; 99US-0132485.
06-MAY-1999; 99US-0132486.
06-MAY-1999; 99US-0132487.
07-MAY-1999; 99US-0132863.
11-MAY-1999; 99US-0134256.
14-MAY-1999; 99US-0134218.
14-MAY-1999; 99US-0134219.
14-MAY-1999; 99US-0134221.
14-MAY-1999; 99US-0134370.
18-MAY-1999; 99US-0134768.
19-MAY-1999; 99US-0134941.
20-MAY-1999; 99US-0135124.
21-MAY-1999; 99US-0135253.
24-MAY-1999; 99US-0135629.
25-MAY-1999; 99US-0136021.
27-MAY-1999; 99US-0136392.
28-MAY-1999; 99US-0136782.
01-JUN-1999; 99US-0137222.
03-JUN-1999; 99US-0137528.
04-JUN-1999; 99US-0137502.
07-JUN-1999; 99US-0137724.
08-JUN-1999; 99US-0138094.
10-JUN-1999; 99US-0138540.
10-JUN-1999; 99US-0138847.
14-JUN-1999; 99US-0139119.
16-JUN-1999; 99US-0139452.
16-JUN-1999; 99US-0139453.
17-JUN-1999; 99US-0139492.
18-JUN-1999; 99US-0139454.
18-JUN-1999; 99US-0139455.
18-JUN-1999; 99US-0139456.
18-JUN-1999; 99US-0139457.
18-JUN-1999; 99US-0139458.
18-JUN-1999; 99US-0139459.
18-JUN-1999; 99US-0139460.
18-JUN-1999; 99US-0139461.
18-JUN-1999; 99US-0139462.
18-JUN-1999; 99US-0139463.
18-JUN-1999; 99US-0139750.
18-JUN-1999; 99US-0139767.
21-JUN-1999; 99US-0139817.
22-JUN-1999; 99US-0139899.
23-JUN-1999; 99US-0140353.
23-JUN-1999; 99US-0140354.
23-JUN-1999; 99US-0140355.
24-JUN-1999; 99US-0140695.
28-JUN-1999; 99US-0140823.
29-JUN-1999; 99US-0140991.
30-JUN-1999; 99US-0141287.
01-JUL-1999; 99US-0141842.
01-JUL-1999; 99US-0142154.
02-JUL-1999; 99US-0142055.
06-JUL-1999; 99US-0142390.
08-JUL-1999; 99US-0142803.
09-JUL-1999; 99US-0142920.
12-JUL-1999; 99US-0142977.
13-JUL-1999; 99US-0143542.
14-JUL-1999; 99US-0143624.
15-JUL-1999; 99US-0144005.
16-JUL-1999; 99US-0144086.
16-JUL-1999; 99US-0144086.
19-JUL-1999; 99US-0144325.
19-JUL-1999; 99US-0144331.
19-JUL-1999; 99US-0144332.
19-JUL-1999; 99US-0144333.
19-JUL-1999; 99US-0144334.
19-JUL-1999; 99US-0144335.
20-JUL-1999; 99US-0144632.
20-JUL-1999; 99US-0144884.
21-JUL-1999; 99US-0144814.
21-JUL-1999; 99US-0145086.
21-JUL-1999; 99US-0145088.
22-JUL-1999; 99US-0145087.
22-JUL-1999; 99US-0145087.
22-JUL-1999; 99US-0145089.
22-JUL-1999; 99US-0145192.
23-JUL-1999; 99US-0145145.
23-JUL-1999; 99US-0145218.
23-JUL-1999; 99US-0145224.
26-JUL-1999; 99US-0145276.
27-JUL-1999; 99US-0145913.
27-JUL-1999; 99US-0145918.
27-JUL-1999; 99US-0145919.
28-JUL-1999; 99US-0145951.
02-AUG-1999; 99US-0146386.
02-AUG-1999; 99US-0146388.
02-AUG-1999; 99US-0146389.
03-AUG-1999; 99US-0147038.
04-AUG-1999; 99US-0147204.
04-AUG-1999; 99US-0147302.
05-AUG-1999; 99US-0147192.
05-AUG-1999; 99US-0147260.
06-AUG-1999; 99US-0147303.
06-AUG-1999; 99US-0147416.
09-AUG-1999; 99US-0147493.
09-AUG-1999; 99US-0147935.
10-AUG-1999; 99US-0148171.
11-AUG-1999; 99US-0148319.
12-AUG-1999; 99US-0148565.
13-AUG-1999; 99US-0148565.
13-AUG-1999; 99US-0148568.
16-AUG-1999; 99US-0149368.
17-AUG-1999; 99US-0149175.
18-AUG-1999; 99US-0149426.
20-AUG-1999; 99US-0149722.
20-AUG-1999; 99US-0149723.
20-AUG-1999; 99US-0149929.
23-AUG-1999; 99US-0149902.
23-AUG-1999; 99US-0149930.
25-AUG-1999; 99US-0150566.
26-AUG-1999; 99US-0150884.
27-AUG-1999; 99US-0151065.
27-AUG-1999; 99US-0151066.
27-AUG-1999; 99US-0151080.
30-AUG-1999; 99US-0151303.
31-AUG-1999; 99US-0151438.
01-SEP-1999; 99US-0151930.

```

Query Match      12.5%  Score 71.5;  DB 21;  Length 175;
Best Local Similarity 28.6%  Pred No. 12;
Matches 26;  Conservative 12;  Mismatches 32;  Indels 21;  Gaps 4
QY      23  KKSDDLDECKKCTCFKTECYIVFEDTVNKECYNVVDSEEL-DQEKVVDENFTENYLT- 80
          |||      |      |      |      |      |      |      |      |
Db      93  KKGS-----KRALPESFFWTPTDAHQE-----DAGDEIHDEVADLIKEDLWMSNPLY 141
QY      81  -----DCEGKDAGNMAAGTGDSEVED 103
          |      |      |      |      |      |      |      |
Db      142 FNNDADEDEDFGDGDGDEGEEDDDDEDERED 172

RESULT_13
AAG17788
ID  AAG17788 standard; Protein; 256 AA.
XX
AC
AAG17788;
XX
DT  17-OCT-2000 (first entry)
XX
DE  Arabidopsis thaliana protein fragment SEQ ID NO: 18942.
XX
XX  Protein identification: signal transduction pathway; metabolic pathway;
KW  hybridisation assay; genetic mapping; gene expression control; promoter;
XX  termination sequence.
XX

```

OS	Arabiopsis thaliana.	
XX	ED1033405-42.	
PN		
XX	06-SEP-2000.	
FD		
XX		
PE	25-FEB-2000;	2000EP-0301439.
XX		
PR	25-FEB-1999;	9905-0121825.
PR	09-MAR-1999;	9905-0123180.
PR	09-MAR-1999;	9905-0123180.
PR	23-MAR-1999;	9905-0125788.
PR	25-MAR-1999;	9905-0126264.
PR	29-MAR-1999;	9905-0126785.
PR	01-APR-1999;	9905-0127462.
PR	06-APR-1999;	9905-0128234.
PR	08-APR-1999;	9905-0128714.
PR	15-APR-1999;	9905-0129845.
PR	19-APR-1999;	9905-0130077.
PR	21-APR-1999;	9905-0130049.
PR	23-APR-1999;	9905-0130510.
PR	23-APR-1999;	9905-0130891.
PR	28-APR-1999;	9905-0131149.
PR	30-APR-1999;	9905-0132048.
PR	30-APR-1999;	9905-0132407.
PR	04-MAY-1999;	9905-0132484.
PR	05-MAY-1999;	9905-0132485.
PR	06-MAY-1999;	9905-0132486.
PR	06-MAY-1999;	9905-0132487.
PR	07-MAY-1999;	9905-0132683.
PR	11-MAY-1999;	9905-0134215.
PR	14-MAY-1999;	9905-0134218.
PR	14-MAY-1999;	9905-0134219.
PR	14-MAY-1999;	9905-0134221.
PR	14-MAY-1999;	9905-0134370.
PR	18-MAY-1999;	9905-0134461.
PR	19-MAY-1999;	9905-0134944.
PR	20-MAY-1999;	9905-0135124.
PR	21-MAY-1999;	9905-0135353.
PR	24-MAY-1999;	9905-0135629.
PR	25-MAY-1999;	9905-0136021.
PR	27-MAY-1999;	9905-0136392.
PR	28-MAY-1999;	9905-0136782.
PR	01-JUN-1999;	9905-0137222.
PR	10-JUN-1999;	9905-0138540.
PR	14-JUN-1999;	9905-0138847.
PR	14-JUN-1999;	9905-0139119.
PR	16-JUN-1999;	9905-0139452.
PR	16-JUN-1999;	9905-0139453.
PR	17-JUN-1999;	9905-0139454.
PR	18-JUN-1999;	9905-0139455.
PR	18-JUN-1999;	9905-0139456.
PR	18-JUN-1999;	9905-0139457.
PR	18-JUN-1999;	9905-0139458.
PR	18-JUN-1999;	9905-0139459.
PR	18-JUN-1999;	9905-0139460.
PR	18-JUN-1999;	9905-0139461.
PR	18-JUN-1999;	9905-0139462.
PR	18-JUN-1999;	9905-0139463.
PR	18-JUN-1999;	9905-0139750.
PR	18-JUN-1999;	9905-0139763.
PR	21-JUN-1999;	9905-0139763.
PR	21-JUN-1999;	9905-0139889.
PR	23-JUN-1999;	9905-0140353.
PR	23-JUN-1999;	9905-0140354.
PR	24-JUN-1999;	9905-0140655.
PR	28-JUN-1999;	9905-0140823.
PR	29-JUN-1999;	9905-0140991.


```

PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143624.
PR 14-JUL-1999; 99US-0144005.
PR 15-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144633.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 27-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0146381.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147034.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147392.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 25-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 27-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 07-SEP-1999; 99US-0151930.
PR 10-SEP-1999; 99US-0152363.
PR 13-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.

```

```

PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155139.
PR 24-SEP-1999; 99US-0155486.
PR 28-SEP-1999; 99US-0155659.
PR 29-SEP-1999; 99US-0156458.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158369.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 25-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 26-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 28-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 12.5%; Score 71.5; DB 21; Length 256;
 Best Local Similarity 28.6%; Pred. No. 19;
 Matches 26; Conservative 12; Mismatches 32; Indels 21; Gaps 4;

```

QY 23 KSSDDECKKTCFTECYIYEDTVNKECYNVVDEGL-DOEKFYVDENFTENYLT-80
DB 174 KGS-----KRALPESFFFTFDQHK-----DAGDEIHDEVADIKEDLWSNPLYT 222
QY 81 -----DOEGKDGAGNAAGTGESDEVED 103
DB 223 FNNDADEDFDGDGDDGDEGEDDDDEED 253

```

```

RESULT 14
AA18263
ID AA18263 standard; Protein: 320 AA.
XX
AC AA18263;
XX
DT 07-NOV-2000 (first entry)
XX
DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:120.
XX
KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
XX
KW antimalarial; malaria; protozoacide; infection; insecticide.
OS Plasmodium falciparum.
XX
PN W0200025728-A2.
XX

```

PD 11-MAY-2000.
 XX
 PF 05-NOV-1999; 99MO-US26796.
 XX
 PR 05-NOV-1998; 98US-0107131.
 XX
 PA (HOFF/) HOFFMAN S.
 PA (CARU/) CARUCCI D.
 PA (GARD/) GARDNER M.
 PA (VENT/) VENTER J C.
 XX
 PI Hoffman S, Carucci D, Gardner M, Venter JC;
 DR WPI: 2000-365347/31.
 XX
 PT Proteins encoded by chromosome 2 of the human malarial parasite,
 PT Plasmodium falciparum, useful as antimalarial vaccines and in the
 PT diagnosis of P.falciparum infection -
 PS Disclosure: Page 283-284; 577pp; English.
 XX
 CC The present invention describes proteins and their fragments (I) encoded
 CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum,
 CC also described are: (1) nucleotide sequences (II) encoding (I); and (2)
 CC vaccines against P. falciparum infection comprising (I) or (II).
 CC (1) and (II) are useful for the development of vaccines against
 CC P. falciparum infection. (1) and polyclonal antisera or a monoclonal
 CC antibody raised to immunogens comprising the sequences of (I), are
 CC useful in the detection of infection with P. falciparum. Furthermore,
 CC (1) (especially when they are rifins or secreted or membrane proteins)
 CC can aid the identification of drugs to treat or prevent P. falciparum
 CC infection, or they can be used to identify drug resistance in
 CC P. falciparum. Sequencing of the plasmodium chromosome 2 and the
 CC subsequent identification of proteins encoded by it will help to expand
 CC our understanding of parasite lifecycle, and provide new targets for
 CC complexity of the parasite lifecycle. Parasite resistance to drugs and mosquito
 CC resistance to insecticides have led to a resurgence of malaria in many
 CC parts of the world, and there is a pressing need for vaccines and new
 CC drugs. AAH70078 to AAH70287 and AAH18144 to AAH18352 represent nucleotide
 CC and protein sequences given in the present invention, but which are not
 CC specifically mentioned within the specification.
 XX
 SQ Sequence 320 AA;
 12.5%; Score 71.5; DB 21; Length 320;
 Query Match Best Local Similarity 27.8%; Pred. No. 25;
 Matches 27; Conservative 9; Mismatches 24; Indels 37; Gaps 5;
 9 EVANRRY-----TDEPKS-----FKKSDLECKR-----TCFKTEYC----- 41
 51 FYALKRYEKSLEKKRDEKSNNDKISIKSYKDFKNELQITIKNEYCICEGIIITNY 110
 42 ---YIVEEDTVNKECYNVVDEELDOEKFEVVDENET 75
 111 DEYIIIEYEMENDSL-----KFEDEVFVLDKNY 140
 DB
 RESULT 15
 ID AAB95086 standard; Protein; 589 AA.
 XX
 AC AAB95086;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human protein sequence SEQ ID NO:16999.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN EP1074617-A2.

XX 07-FEB-2001.
 PD
 XX 28-JUL-2000; 2000EP-0116126.
 XX
 PF 29-JUL-1999; 99JP-0248036.
 XX 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakematsu A, Nagai K, Otsuki T;
 DR WPI: 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 PS Claim 8; SEQ ID 16999; 2537pp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 SQ Sequence 589 AA;
 12.4%; Score 71; DB 22; Length 589;
 Query Match Best Local Similarity 23.8%; Pred. No. 59;
 Matches 30; Conservative 23; Mismatches 37; Indels 36; Gaps 6;
 12 NKRITDDEKSEKFS-----SDLECKRY-----CFKTE---ICYIVE 46
 158 SKEFTQNKREKKNIVYHTDSSLEKORTLDSGTSEIYKTPRICSTREKMSVQLI 217
 47 DTVNKECYNVVDEELDOEKFEVVD-----EKFTEYTLDECGKADAGNAGTGDES 97
 218 MTRSDGVEYENSTDEGMDCKDALEDESEYSEIGSESENETSY-GRASGDGDSSEDE 276
 98 DEVEDD 103
 277 EE-DED 281
 DB

Search completed: February 25, 2003, 02:32:13
 Job time : 28 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 25, 2003, 02:31:10 ; Search time 15 Seconds
(without alignments)
202.037 Million cell updates/sec

Title: US-09-936-737a-2
Perfect score: 573
Sequence: 1 EBERDCWTFYANRKYTFDK.....GKDAGNAGTDESDEVED 103

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/PCITUS.COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	80.5	14.0	126	1	US-08-044-547-3
2	80.5	14.0	147	1	US-08-044-547-1
3	80.5	14.0	412	2	US-08-741-134-2
4	75.5	13.2	172	4	US-09-187-789-16
5	75.5	13.2	172	4	US-09-139-600-11
6	75.5	13.2	418	3	US-08-258-287B-58
7	75.5	13.2	418	3	US-08-368-704C-56
8	75.5	13.2	419	3	US-08-258-287B-57
9	75.5	13.2	419	3	US-08-368-704C-55
10	69.5	12.1	124	1	US-08-170-360-1
11	69.5	12.1	674	4	US-08-961-083-200
12	67	11.7	124	1	US-08-170-360-2
13	67	11.7	124	2	US-08-888-497-41
14	67	11.7	124	4	US-09-362-230-41
15	67	11.7	124	5	PCT-US94-07926-41
16	67	11.7	146	4	US-08-888-497-34
17	67	11.7	146	4	US-09-362-230-34
18	67	11.7	146	4	PCT-US94-07926-34
19	67	11.7	299	4	US-09-069-023-20
20	66.5	11.6	86	4	US-09-134-001C-3986
21	66.5	11.6	2182	2	US-08-487-826B-16
22	65.5	11.3	725	1	US-08-170-360-3
23	64.5	11.3	126	2	US-08-888-497-38
24	64.5	11.3	126	4	US-09-362-230-38
25	64.5	11.3	126	5	PCT-US94-07926-38
26	64.5	11.3	148	2	US-08-888-497-36
27	64.5	11.3			

28	64.5	11.3	148	4	US-09-362-230-36	Sequence 36, Appl
29	64.5	11.3	148	5	PCT-US94-07926-36	Sequence 36, Appl
30	64	11.2	519	4	US-09-172-841-55	Sequence 55, Appl
31	63	11.0	294	4	US-09-355-160D-8	Sequence 8, Appl
32	63	11.0	1658	2	US-08-609-049A-13	Sequence 13, Appl
33	63	11.0	1658	4	US-09-170-996-13	Sequence 13, Appl
34	63	11.0	1726	2	US-08-609-049A-30	Sequence 30, Appl
35	63	11.0	1726	4	US-09-170-996-30	Sequence 30, Appl
36	62.5	10.9	250	2	US-08-861-269-5	Sequence 5, Appl
37	62.5	10.9	250	2	US-09-134-596-5	Sequence 5, Appl
38	62.5	10.9	250	3	US-09-293-273-5	Sequence 19, Appl
39	62.5	10.9	1172	1	US-08-313-288B-19	Sequence 36, Appl
40	62.5	10.9	2555	3	US-09-058-489-36	Sequence 4, Appl
41	61.5	10.7	346	2	US-08-702-153-2	Sequence 3, Appl
42	61.5	10.7	346	2	US-08-702-153-4	Sequence 3, Appl
43	61.5	10.7	416	2	US-09-211-930-3	Sequence 3, Appl
44	61.5	10.7	416	3	US-09-340-993-3	Sequence 3, Appl
45	61.5	10.7	416	4	US-09-468-442-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1
US-08-044-547-3
Sequence 3, Application US/08044547
Patent No. 5324715
GENERAL INFORMATION:
APPLICANT: Connolly, Thomas M.
TITLE OF INVENTION: Protein for Inhibiting
Collagen-Stimulated Platelet Aggregation
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merck & Co., Inc.
STREET: P.O. Box 2000
CITY: Rahway
STATE: New Jersey
COUNTRY: US
ZIP: 07065
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/044,547
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/594,917
FILING DATE: 09-OCT-1990
ATTORNEY/AGENT INFORMATION:
NAME: Parr, Richard S.
REGISTRATION NUMBER: 32,586
REFERENCE/DOCKET NUMBER: 18053
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 594-4958
TELEFAX: (908) 594-4720
TELEX: 138825
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 126 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-044-547-3
Query Match 14.0%; Score 80.5; DB 1; Length 126;
Best Local Similarity 32.7%; Pred. No. 0.21;
Matches 18; Conservative 10; Mismatches 24; Indels 3; Gaps 1;

OY 5 DCWTFYANKRYTDFDKSFKSSDDECKKTKCTEY---CYIVEDIYVNECYIN 56
 DB 45 DCWSKRPCKWKLDPNLLTKTEFTSVDECRKMCESAVESCYILQINTEFTNECYRN 99

RESULT 2
 US-08-044-547-1
 ; Sequence 1, Application US/08044547
 ; Patent No. 5324715
 ; GENERAL INFORMATION:
 ; APPLICANT: Connolly, Thomas M.
 ; APPLICANT: Keller, Paul M.
 ; TITLE OF INVENTION: Protein for Inhibiting
 ; TITLE OF INVENTION: Collagen-Stimulated Platelet Aggregation
 ; NUMBER OF SEQUENCES: 12
 ; CORRESPONDENCE ADDRESSES:
 ; ADDRESSEE: Merck & Co., Inc.
 ; STREET: P.O. Box 2000
 ; CITY: Rahway
 ; STATE: New Jersey
 ; COUNTRY: US
 ; ZIP: 07065
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/044,547
 ; FILING DATE:
 ; CLASSIFICATION: 514
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/594,917
 ; FILING DATE: 09-OCT-1990
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Parr, Richard S.
 ; REGISTRATION NUMBER: 32,586
 ; REFERENCE/DOCKET NUMBER: 18053
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (908) 594-4958
 ; TELEFAX: (908) 594-4720
 ; TELEX: 138825
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 147 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: unknown
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: protein
 ; US-08-044-547-1

Query Match 14.0%; Score 80.5; DB 1; Length 147;
 Best Local Similarity 32.7%; Pred. No. 0.25; Mismatches 24; Indels 3; Gaps 1;
 Matches 18; Conservative 10; Mismatches 24; Indels 3; Gaps 1;
 OY 5 DCWTFYANKRYTDFDKSFKSSDDECKKTKCTEY---CYIVEDIYVNECYIN 56
 DB 66 DCWSKRPCKWKLDPNLLTKTEFTSVDECRKMCESAVESCYILQINTEFTNECYRN 120

RESULT 3
 US-08-741-134-2
 ; Sequence 2, Application US/08741134
 ; Patent No. 5861498
 ; GENERAL INFORMATION:
 ; APPLICANT: Litwack, Gerald
 ; APPLICANT: Alnemri, Emad S.
 ; APPLICANT: Fernandez-Alnemri, Teresa
 ; TITLE OF INVENTION: IMMUNOPHILIN FKBP46 AND COMPOSITIONS FOR MAKING
 ; TITLE OF INVENTION: AND
 ; TITLE OF INVENTION: METHODS OF USING THE SAME
 ; NUMBER OF SEQUENCES: 6
 ; CORRESPONDENCE ADDRESS:

ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5861498ris
 STREET: One Liberty Place - 46th floor
 CITY: Philadelphia
 STATE: Pennsylvania
 COUNTRY: USA
 ZIP: 19103
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: Windows 3.11
 ; SOFTWARE: WordPerfect for Windows 6.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/741,134
 ; FILING DATE:
 ; CLASSIFICATION: 424
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 60/007,163
 ; FILING DATE: 01-NOV-1995
 ; CLASSIFICATION: 424
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Deluca, Mark
 ; REGISTRATION NUMBER: 33,229
 ; REFERENCE/DOCKET NUMBER: TJU-2090
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 215-568-3100
 ; TELEFAX: 215-568-3439
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 412 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-741-134-2

Query Match 14.0%; Score 80.5; DB 2; Length 412;
 Best Local Similarity 30.6%; Pred. No. 0.87; Mismatches 29; Indels 27; Gaps 5;
 Matches 30; Conservative 12; Mismatches 29; Indels 27; Gaps 5;
 OY 11 ANRRYTFEDKSFKRS---SDIDECKKTKCTEYCYIVEDIYVNECYINVDGEELDO 65
 DB 133 ANKRAKPPDKKAKNSAPAAESDSDDDD-----EDQLK-----FLDGEDIDT 174

OY 66 EKPVVDENFTENTYLTDEGKDNAGNAAGTGPDESDEVED 103
 DB 175 DE--NDESFKRN--TSABGDSDSEDEDEDEDEDD 208

RESULT 4
 US-09-187-789-16
 ; Sequence 16, Application US/09187789
 ; Patent No. 6340740
 ; GENERAL INFORMATION:
 ; APPLICANT: Alnemri, Emad S.
 ; APPLICANT: Fernandez-Alnemri, Teresa
 ; TITLE OF INVENTION: CASPASE-14, AN APOPTOTIC PROTEASE, NUCLEIC ACID ENCODING
 ; TITLE OF INVENTION: AND METHODS OF USE
 ; FILE REFERENCE: 480140.434C1
 ; CURRENT APPLICATION NUMBER: US/09/187,789
 ; CURRENT FILING DATE: 1998-11-06
 ; NUMBER OF SEQ ID NOS: 78
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 16
 ; LENGTH: 172
 ; TYPE: PRT
 ; ORGANISM: Mus musculus
 ; US-09-187-789-16

Query Match 13.2%; Score 75.5; DB 4; Length 172;
 Best Local Similarity 35.6%; Pred. No. 1; Mismatches 14; Indels 19; Gaps 6;
 Matches 26; Conservative 14; Mismatches 14; Indels 19; Gaps 6;
 OY 44 VEDIYVNECYINVDGEEL---DQKRYVD--ENFTENTYLTDEGK--DAGNAGTG 94

Wed Feb 26 15:59:00 2003

us-09-936-737a-2.ra1

Page 3

Db 26 VFDDLVK-----NVLNDELKIGESASFLINKAENLVENLEKTDMAKIFAGHANSQ 81
QY 95 DE-----SDEVD 101
82 EQLSLQFSNDEDD 94

RESULT 5
US-09-139-600-11
Sequence 11, Application US/09139600
Patent No. 6432628

GENERAL INFORMATION:
APPLICANT: Alnemir, Emad S.
TITLE OF INVENTION: Fernandez-Alnemir, Teresa
TITLE OF INVENTION: CASPASE-14, AN APOPTOTIC PROTEASE, NUCLEIC ACID ENCODING
FILE REFERENCE: 480140.434
CURRENT APPLICATION NUMBER: US/09/139,600
CURRENT FILING DATE: 1998-08-25
NUMBER OF SEQ ID NOS: 65
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 11
LENGTH: 172
TYPE: PRT
ORGANISM: Mus musculus
US-09-139-600-11

Query Match 13.2%; Score 75.5; DB 4; Length 172;
Best Local Similarity 35.6%; Pred. No. 1;
Matches 26; Conservative 14; Mismatches 14; Indels 19; Gaps 6;

QY 44 VFEDYVNECEYVNVGDEEL---DQKRVVD--ENFTENYL--TDCEGK-DAGNAGTG 94
26 VFDDLVK-----NVLNDELKIGESASFLINKAENLVENLEKTDMAKIFAGHANSQ 81

QY 95 DE-----SDEVD 101
82 EQLSLQFSNDEDD 94

RESULT 6
US-08-258-287B-58
Sequence 58, Application US/08258287B
Patent No. 6083735
GENERAL INFORMATION:
APPLICANT: Yuan, Junyong
TITLE OF INVENTION: Programmed Cell Death Genes and Proteins
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/258,287B
FILING DATE: 10-JUN-1994
CLASSIFICATION: 435
PRIOR APPLICATION NUMBER: US 08/080,850
APPLICATION NUMBER: 24-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Bugalsky, Lawrence B.
REGISTRATION NUMBER: 35,086
REFERENCE/DOCKET NUMBER: 0609.3920001
TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
TELEX: 24636 SSK
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 418 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-258-287B-58

Query Match 13.2%; Score 75.5; DB 3; Length 418;
Best Local Similarity 35.6%; Pred. No. 3;
Matches 26; Conservative 14; Mismatches 14; Indels 19; Gaps 6;

QY 44 VFEDYVNECEYVNVGDEEL---DQKRVVD--ENFTENYL--TDCEGK-DAGNAGTG 94
26 VFDDLVK-----NVLNDELKIGESASFLINKAENLVENLEKTDMAKIFAGHANSQ 81
QY 95 DE-----SDEVD 101
82 EQLSLQFSNDEDD 94

RESULT 7
US-08-368-704C-56
Sequence 56, Application US/08368704C
Patent No. 6087160
GENERAL INFORMATION:
APPLICANT: Yuan, Junyong
TITLE OF INVENTION: Programmed Cell Death Genes and Proteins
NUMBER OF SEQUENCES: 95
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/368,704C
FILING DATE: 4-JAN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/258,287
FILING DATE: 10-JUN-1994
CLASSIFICATION: 435
PRIOR APPLICATION NUMBER: US 08/080,850
APPLICATION NUMBER: 24-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Bugalsky, Lawrence B.
REGISTRATION NUMBER: 35,086
REFERENCE/DOCKET NUMBER: 0609.3920002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 418 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-368-704C-56

Query Match 13.2%; Score 75.5; DB 3; Length 419;
Best Local Similarity 35.6%; Pred. No. 3;
Matches 26; Conservative 14; Mismatches 14; Indels 19; Gaps 6;

QY 44 VFEDTVNKECYNVVDGEEL---DQEKFYVD--ENFTENTL--TDEGK-DAGNAGTG 94
DB 26 VFDDLVK-----NVLNGDELKIGESASFIINKAKENLVENFLEKTDMAKIFAGHTANSQ 94
QY 95 DE-----SDEVD 101
DB 82 EQLSLOFSNDEDD 94

RESULT 8

US-08-258-287B-57
; Sequence 57, Application US/08258287B
; Patent No. 6083735
; GENERAL INFORMATION:
; APPLICANT: Yuan, Junying
; APPLICANT: Miura, Masayuki
; TITLE OF INVENTION: Programmed Cell Death Genes and Proteins
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 10-JUN-1994
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/080,850
; FILING DATE: 24-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Bugalsky, Lawrence B.
; REGISTRATION NUMBER: 35,086
; REFERENCE/DOCKET NUMBER: 0609.3920001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 419 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-258-287B-57

Query Match 13.2%; Score 75.5; DB 3; Length 419;
Best Local Similarity 35.6%; Pred. No. 3;
Matches 26; Conservative 14; Mismatches 14; Indels 19; Gaps 6;

QY 44 VFEDTVNKECYNVVDGEEL---DQEKFYVD--ENFTENTL--TDEGK-DAGNAGTG 94
DB 26 VFDDLVK-----NVLNGDELKIGESASFIINKAKENLVENFLEKTDMAKIFAGHTANSQ 94
QY 95 DE-----SDEVD 101
DB 82 EQLSLOFSNDEDD 94

RESULT 9
US-08-368-704C-55

; Sequence 55, Application US/08368704C
; Patent No. 6087160
; GENERAL INFORMATION:
; APPLICANT: Yuan, Junying
; APPLICANT: Miura, Masayuki
; TITLE OF INVENTION: Programmed Cell Death Genes and Proteins
; NUMBER OF SEQUENCES: 95
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 4-JAN-1995
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/258,287
; FILING DATE: 10-JUN-1994
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/080,850
; FILING DATE: 24-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Bugalsky, Lawrence B.
; REGISTRATION NUMBER: 35,086
; REFERENCE/DOCKET NUMBER: 0609.3920002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 55:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 419 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-368-704C-55

Query Match 13.2%; Score 75.5; DB 3; Length 419;
Best Local Similarity 35.6%; Pred. No. 3;
Matches 26; Conservative 14; Mismatches 14; Indels 19; Gaps 6;

QY 44 VFEDTVNKECYNVVDGEEL---DQEKFYVD--ENFTENTL--TDEGK-DAGNAGTG 94
DB 26 VFDDLVK-----NVLNGDELKIGESASFIINKAKENLVENFLEKTDMAKIFAGHTANSQ 81
QY 95 DE-----SDEVD 101
DB 82 EQLSLOFSNDEDD 94

RESULT 10

US-08-170-360-1
; Sequence 1, Application US/08170360
; Patent No. 5656602
; GENERAL INFORMATION:
; APPLICANT: Tseng, Albert P. S.
; APPLICANT: Ingalls, Adam
; TITLE OF INVENTION: PL2 INHIBITORY COMPOUNDS
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rothwell, Figg Ernst & Kurz
; STREET: Suite 701-E, 555 Thirteenth St., N.W
; CITY: Washington
; STATE: D. C.

COUNTRY: U.S.A.
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/170,360
FILING DATE: 03-MAR-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PCT/AU92/00333
FILING DATE: 06-JUL-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PK 7058
FILING DATE: 04-JUL-1991
ATTORNEY/AGENT INFORMATION:
NAME: Ernst, Barbara G.
REGISTRATION NUMBER: 30,377
REFERENCE/DOCKET NUMBER: 1871-104A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)783-6040
TELEFAX: (202)783-6031
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 124 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
US-08-170-360-1

Query Match 12.1%; Score 69.5; DB 1; Length 124;
Best Local Similarity 27.1%; Pred. No. 2.9;
Matches 16; Conservative 8; Mismatches 20; Indels 15; Gaps 1;

QY 28 LDECKTKCTKCTCYCTVDETVNKECYVNVVGEELDEKFPVDENFENTLTDCCKD 86
DB 38 VDELRCETHDNCY-----RDAKNLDSCKTLVNPITESTISCSNTE 81

RESULT 11
US-08-961-083-200
Sequence 200, Application US/08961083
Patent No. 6159469
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:

NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 200:
SEQUENCE CHARACTERISTICS:
LENGTH: 674 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-961-083-200

Query Match 12.1%; Score 69.5; DB 4; Length 674;
Best Local Similarity 23.6%; Pred. No. 23;
Matches 25; Conservative 17; Mismatches 35; Indels 29; Gaps 5;

QY 7 WTFYANKRYTDEPKSFKSSDLDECKT-----CFTEYCYTFEDTVNKECYNVVGE 61
DB 322 WIKQNKRLKRPD-----SEILNCRYTSBQJSCYKSDISYTF-----LRYFIADFV 368

QY 62 ELDEKEF-----VDENFENTENLTDCG-----KDGMAAGTGD 96
DB 369 QEDKALYLDCCDLVTKNLDLFFATDLQDYPPLAAYRDFGGRAYFGOE 414

RESULT 12
US-08-170-360-2
Sequence 2, Application US/08170360
Patent No. 5656602
GENERAL INFORMATION:
APPLICANT: Tseng, Albert P. S.
APPLICANT: Inglis, Adam
TITLE OF INVENTION: PLAZ INHIBITORY COMPOUNDS
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Rothwell, Figg Ernst & Kurz
STREET: Suite 701-E, 555 Thirteenth St., N.W
CITY: Washington
STATE: D. C.
COUNTRY: U.S.A.
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/170,360
FILING DATE: 03-MAR-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PCT/AU92/00333
FILING DATE: 06-JUL-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PK 7058
FILING DATE: 04-JUL-1991
ATTORNEY/AGENT INFORMATION:
NAME: Ernst, Barbara G.
REGISTRATION NUMBER: 30,377
REFERENCE/DOCKET NUMBER: 1871-104A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)783-6040
TELEFAX: (202)783-6031
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 124 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein

HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
US-08-170-360-2

Query Match 11.7%; Score 67; DB 1; Length 124;
Best Local Similarity 22.7%; Pred. No. 5.4;
Matches 17; Conservative 14; Mismatches 22; Indels 22; Gaps 3;

Query 27 DLDECKKTCFTEKYIVFEDTVNKECYNNVDSGEILDQKRFVVDENFTENYLTDCGKD 86
Db 40 DLDRG---CQTHHCY-----NQAKKLESCFLIDNPYNTYISYK----- 77

QY 87 AGNAGTGSDEVD 101
DB 78 SGNVITCSDKNDCE 92

RESULT 13
US-08-888-497-41
Sequence 41, Application US/08888497
Patent No. 5972677

GENERAL INFORMATION:

APPLICANT: Tischfield, Jay A.
TITLE OF INVENTION: Mammalian Phospholipase A2 Nucleotide
Sequences and Low Molecular Weight Amino Acid Sequences
TITLE OF INVENTION: Encoded Thereby, Antisense Sequences and Nucleotide
Sequences Having Internal Ribosome Binding Sites
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Ruden, Barnett, McClosky, Smith, Schuster &
STREET: 200 East Broward Boulevard
CITY: Fort Lauderdale
STATE: FL
COUNTRY: USA
ZIP: 33301

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/888,497
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/651,405
FILING DATE:
APPLICATION NUMBER: US 08/097,354
FILING DATE: 26-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: Manso, Peter J.
REGISTRATION NUMBER: 32,264
REFERENCE/DOCKET NUMBER: IN21044-5
TELECOMMUNICATION INFORMATION:
TELEPHONE: 305-527-2498
TELEFAX: 305-764-4996

INFORMATION FOR SEQ ID NO: 41:
SEQUENCE CHARACTERISTICS:
LENGTH: 124 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-888-497-41

Query Match 11.7%; Score 67; DB 2; Length 124;
Best Local Similarity 22.7%; Pred. No. 5.4;
Matches 17; Conservative 14; Mismatches 22; Indels 22; Gaps 3;

QY 27 DLDECKKTCFTEKYIVFEDTVNKECYNNVDSGEILDQKRFVVDENFTENYLTDCGKD 86

Db 40 DLDRG---CQTHHCY-----NQAKKLESCFLIDNPYNTYISYK----- 77
QY 87 AGNAGTGSDEVD 101
DB 78 SGNVITCSDKNDCE 92

RESULT 14
US-09-362-230-41
Sequence 41, Application US/09362230
Patent No. 6352849

GENERAL INFORMATION:
APPLICANT: Tischfield, Jay A.
TITLE OF INVENTION: Mammalian Phospholipase A2 Nucleotide
Sequences and Low Molecular Weight Amino Acid Sequences
TITLE OF INVENTION: Encoded Thereby, Antisense Sequences and Nucleotide
Sequences Having Internal Ribosome Binding Sites
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Ruden, Barnett, McClosky, Smith, Schuster &
STREET: 200 East Broward Boulevard
CITY: Fort Lauderdale
STATE: FL
COUNTRY: USA
ZIP: 33301

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/362,230
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/888,497
FILING DATE:
APPLICATION NUMBER: US 08/097,354
FILING DATE: 26-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: Manso, Peter J.
REGISTRATION NUMBER: 32,264
REFERENCE/DOCKET NUMBER: IN21044-5
TELECOMMUNICATION INFORMATION:
TELEPHONE: 305-527-2498
TELEFAX: 305-764-4996

INFORMATION FOR SEQ ID NO: 41:
SEQUENCE CHARACTERISTICS:
LENGTH: 124 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-362-230-41

Query Match 11.7%; Score 67; DB 4; Length 124;
Best Local Similarity 22.7%; Pred. No. 5.4;
Matches 17; Conservative 14; Mismatches 22; Indels 22; Gaps 3;

QY 27 DLDECKKTCFTEKYIVFEDTVNKECYNNVDSGEILDQKRFVVDENFTENYLTDCGKD 86
Db 40 DLDRG---CQTHHCY-----NQAKKLESCFLIDNPYNTYISYK----- 77
QY 87 AGNAGTGSDEVD 101
DB 78 SGNVITCSDKNDCE 92

RESULT 15
PCT-US94-07926-41


```

: Sequence 41, Application PC/TUS9407926
: GENERAL INFORMATION:
: APPLICANT: Tischfield, Jay A.
: TITLE OF INVENTION: Mammalian Phospholipase A2 Nucleotide
: TITLE OF INVENTION: Sequences and Low Molecular Weight Amino Acid Sequences
: TITLE OF INVENTION: Encoded Thereby, Antisense Sequences and Nucleotide
: NUMBER OF INVENTION: Sequences Having Internal Ribosome Binding Sites
: CORRESPONDENCE ADDRESSES: 44
: ADDRESSEE: Ruden, Barnett, McClosky, Smith, Schuster &
: STREET: 200 East Broward Boulevard
: CITY: Fort Lauderdale
: STATE: FL
: COUNTRY: USA
: ZIP: 33301
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US94/07926
: FILING DATE: 15-JUL-1994
: CLASSIFICATION:
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: US 08/097,354
: FILING DATE: 26-JUL-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Manso, Peter J.
: REGISTRATION NUMBER: 32,264
: REFERENCE/DOCKET NUMBER: IN21044-5
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 305-527-2498
: TELEFAX: 305-764-4996
: INFORMATION FOR SEQ ID NO: 41:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 124 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: PCT-US94-07926-41

Query Match      11.7%; Score 67; DB 5; Length 124;
Best Local Similarity 22.7%; Pred. NO. 5.4;
Matches 17; Conservative 14; Mismatches 22; Indels 22; Gaps 3;

QY 27 DLDECKKTCFTEYCYIVFEDTVNKECYVNVYDGEELDQERFVVDENFTENYLTDCBGKD 86
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 40 DLDRG---CQTHDRY-----NOAKKLESCKFLIDNPFITNTYSRC----- 77
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 87 ACNAGTGDSEVD 101
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 78 SCNVITCSDKNNDC 92

```

Search completed: February 25, 2003, 02:33:47
 Job time : 16 secs

APPLICANT: Mohamath, Raodoh

```

; APPLICANT: Indrias, Carol Y.
; APPLICANT: Pan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.512
; CURRENT APPLICATION NUMBER: US/09/833,790
; CURRENT FILING DATE: 2001-04-11
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 252
; LENGTH: 156
; TYPE: PRF
; ORGANISM: Homo sapiens
US-09-833-790-252

```

```

Query Match      12.3%; Score 70.5; DB 10; Length 156;
Best Local Similarity 21.8%; Pred. No. 2.1;
Matches 26; Conservative 20; Mismatches 50; Indels 23; Gaps 6;

```

```

QY 1 EERE-----DCWTFYANKRYTDFDKS-----FKKSSD-----LDECKKTCFTEYCYIV 44
DB 30 EDEKLLKSKCTLYVGLSFTTEQYELFSKSGDIKTIIMGLDKKKTAA--CGFCFYE 87
DB 88 YYSRADNEMNARYINGTRLDLR--IIFDMADGPK--EGRGYGRSGGVAVDEYRQD 141

```

```

RESULT 3
US-09-815-242-11522
; Sequence 11522, Application US/09815242
; Patent No. US2002061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11522
; LENGTH: 1167
; TYPE: PRF
; ORGANISM: Helicobacter pylori
US-09-815-242-11522

```

```

Query Match      12.2%; Score 70; DB 10; Length 1167;
Best Local Similarity 30.8%; Pred. No. 23;
Matches 24; Conservative 12; Mismatches 24; Indels 18; Gaps 4;
QY 19 DKFRKSSDLDECKKTCFTEYCYIVFEDTVNKECYIVNVGDELDEKTFVDENFTENY 78

```

```

DB 240 DKTEQSSNDPKKAQKTLFDA-----LILQDLAN--AVYVMPK-----LGRNRYWENF 286
QY 79 LTDCEKADGNAAGTIDE 96
DB 287 -----AKRTGNARTLINE 299

```

```

RESULT 4
US-09-765-272-200
; Sequence 200, Application US/09765272
; Patent No. US2002061545A1
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8512
; TELEFAX: (301) 309-8504
; INFORMATION FOR SEQ ID NO: 200:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 674 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 200:
US-09-765-272-200

```

```

Query Match      12.1%; Score 69.5; DB 10; Length 674;
Best Local Similarity 23.6%; Pred. No. 14;
Matches 25; Conservative 17; Mismatches 35; Indels 29; Gaps 5;
QY 7 WTFYANKRYTDFDKSKSSDLDECKKT-----CEKTEYCYIVFEDTVNKECYIVNVGDE 61
DB 322 WIKQUNKRLEKFD-----SEIINCRTVSEQISCKSDISYVF-----LRYFIADFV 368
QY 62 ELDDEKF-----VDENFTENYLTDCG-----KDAGNAAGTIDE 96
DB 369 QEDKALYLDCLVVTKNLDDLFRATDLDYPIAAVDFGGRAYGGE 414

```

```

RESULT 5
US-09-993-999-7
; Sequence 7, Application US/09993999
; Patent No. US20020110891A1
; GENERAL INFORMATION:
; APPLICANT: Ho, I-Cheng
; APPLICANT: Arm, Jonathan P.
; APPLICANT: Austen, K. Frank

```

```

; APPLICANT: Glimcher, Laurie H.
; TITLE OF INVENTION: Phospholipase A2 Group Preferentially
; FILE REFERENCE: HUI-046
; CURRENT APPLICATION NUMBER: US/09/993,999
; PRIOR FILING DATE: 2001-11-06
; PRIOR APPLICATION NUMBER: 60/246,316
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 146
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-993-999-7

```

```

Query Match
Best Local Similarity 11.9%; Score 68; DB 10; Length 146;
Matches 15; Conservative 9; Mismatches 18; Indels 18; Gaps 2;

```

```

QY 27 DLDECKKTCFTEKCYIVFEDTVNKECYVNVVDGEELDOEKFYVDENFTENYLTDCCKD 86
DB 62 DLDRG---CQTHRCY-----SOAKLSECKFLDNPFTNTSYSCSGSE 103

```

```

RESULT 6
US-09-801-574-57
; Sequence 57, Application US/09801574
; Patent No. US20020081592A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Peijiang Jeremy
; TITLE OF INVENTION: Reproduction-Specific Genes
; FILE REFERENCE: 0399.2007-002
; CURRENT APPLICATION NUMBER: US/09/801,574
; PRIOR FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: 60/187,518
; PRIOR FILING DATE: 2000-03-07
; PRIOR APPLICATION NUMBER: 60/261,557
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 57
; LENGTH: 2789
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-801-574-57

```

```

Query Match
Best Local Similarity 11.6%; Score 66.5; DB 10; Length 2789;
Matches 25; Conservative 13; Mismatches 23; Indels 35; Gaps 6;

```

```

QY 11 ANRKYTDPRK-----SPKSSDLDECKKTC-----FKTECYIVFEDTVNKECYVNVVDG 60
DB 1349 SSKSYLDRKRLVYDSFMASTVPHCEOSCKREKELKTEQC-----SSGNCILH--TDG 1399
QY 61 EELDOEKFYVDENFTENYLTDCGKDGAGNAAGTDE 96
DB 1400 NE-----TWTENYELDV-----ASTEDD 1419

```

```

RESULT 7
US-09-974-298-114
; Sequence 114, Application US/09974298
; Patent No. US20020156263A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Huel-Mei
; TITLE OF INVENTION: GENES EXPRESSED IN BREAST CANCER
; FILE REFERENCE: PA-0037 P
; CURRENT APPLICATION NUMBER: US/09/974,298
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 60/238,331
; PRIOR FILING DATE: 2000-05-10

```

```

; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 114
; LENGTH: 1125
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Inocyte ID No. US20020156263A1 3267030C01
US-09-974-298-114

```

```

Query Match
Best Local Similarity 11.3%; Score 65; DB 9; Length 1125;
Matches 24; Conservative 24; Mismatches 39; Indels 18; Gaps 4;

```

```

QY 1 EEREDCWTFVANKRYTDPDSSFKSSDLDECKKTCFTEKCYIVFEDTVNKECYVNVVDG 60
DB 1032 EEDD-----SEKEBEDKEMELQEEKCEKPGQDEE-----EEEEEE-----VEE 1075
QY 61 EELDOEKFYVDENFTENYLTDCGKDGAGNAAG--TGDESDVDED 103
DB 1076 EEVEEAENEGEAEKTEGIMKMDRAESQASLSGQKVESESDVSE 1120

```

```

RESULT 8
US-09-925-297-511
; Sequence 511, Application US/09925297
; Patent No. US20020081659A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA105
; CURRENT APPLICATION NUMBER: US/09/925,297
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; NUMBER OF SEQ ID NOS: 928
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 511
; LENGTH: 156
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (156)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-297-511

```

```

Query Match
Best Local Similarity 11.3%; Score 64.5; DB 10; Length 156;
Matches 17; Conservative 5; Mismatches 20; Indels 15; Gaps 2;

```

```

QY 28 LDECKKTCFTEKCYIVFEDTVNKECYVNVVDGEELDOEKFYVDENFTENYLTDCG 84
DB 68 VDELDRKCCQTHNCR-----DQAKK-----LDSCFKFLDNPFTHTYSGSG 109

```

```

RESULT 9
US-10-042-417-52
; Sequence 52, Application US/10042417
; Patent No. US20020123082A1
; GENERAL INFORMATION:
; APPLICANT: Pagano, M.
; TITLE OF INVENTION: METHODS TO IDENTIFY COMPOUNDS USEFUL FOR THE TREATMENT OF
; FILE REFERENCE: 5914-090-999
; CURRENT APPLICATION NUMBER: US/10/042,417
; PRIOR FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: 60/260,179
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 89

```

Wed Feb 26 15:59:01 2003

us-09-936-737a-2.rapb

Page 4

```
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 52
; LENGTH: 590
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: all Xaa positions
; OTHER INFORMATION: xaa-unknown amino acid residue
US-10-042-417-52

Query Match
Best Local Similarity 27.1%; Score 64; DB 12; Length 590;
Matches 23; Conservative 14; Mismatches 38; Indels 10; Gaps 4;

QY 16 TDFPKSKKSSDDECKKTCFTECYIVFEDTVNKECYIVD---GEEIDQ---KP 68
Db 236 TSHSGFLKTS-TSKITSTAMKMK--DITMSTQYACLDLTKNGIGEEIDNHPWTKP 312
QY 69 VVDENFTENYITDCEGKDAGNAGT 93
Db 313 VSENFETSPYWMLEADLADIEDT 337

RESULT 10
US-10-092-219-8
; Sequence 8, Application US/10092219
; Patent No. US2002015114A1
; GENERAL INFORMATION:
; APPLICANT: Domio, Jan
; TITLE OF INVENTION: No. US2002015114A1el Lipid Kinase
; FILE REFERENCE: 1064HG/50947
; CURRENT APPLICATION NUMBER: US/10/092,219
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: PCT/GB98/00244
; PRIOR FILING DATE: 1998-01-27
; PRIOR APPLICATION NUMBER: 09/355,160
; PRIOR FILING DATE: 1998-10-01
; PRIOR APPLICATION NUMBER: 9701653.1
; PRIOR FILING DATE: 1997-01-28
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 294
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-092-219-8

Query Match
Best Local Similarity 11.0%; Score 63; DB 12; Length 294;
Matches 19; Conservative 18; Mismatches 25; Indels 4; Gaps 2;

QY 12 NRRYDFDKSKSSDDECKKTCFTECYIVFEDTVNKECYIVDDEEIDQ---EKF 68
Db 45 DRWMDSPRGFELSSSTQROGFNKQDYDMVFPFLDSOKRAVD-IDVEKLTQALEKI 103
QY 69 VVDENF 74
Db 104 LLDDNF 109

RESULT 11
US-09-930-871-8
; Sequence 8, Application US/09930871
; Patent No. US20020076780A1
; GENERAL INFORMATION:
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Mathur, Daniel
; APPLICANT: Mathur, Brian
; TITLE OF INVENTION: No. US20020076780A1el Human Ion Channel Proteins and Polynucleo
; TITLE OF INVENTION: Same
; FILE REFERENCE: LEX-0216-USA
; CURRENT APPLICATION NUMBER: US/09/930,871
```

```
; CURRENT FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: US 60/225,989
; PRIOR FILING DATE: 2000-08-16
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 1381
; TYPE: PRT
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(1381)
; OTHER INFORMATION: xaa = Any Amino Acid
US-09-930-871-8

Query Match
Best Local Similarity 11.0%; Score 63; DB 10; Length 1381;
Matches 28; Conservative 10; Mismatches 31; Indels 30; Gaps 6;

QY 10 YANRKYTDF-DKSF-KKSSDDECKKTCFTECYIVFEDTVNKE--CYIVVD--GEEI 63
Db 1014 YVKKRIYEFIQOSFTRKQKILDEIKP-----LDDLNNKSCMSNHTXETGKDL 1062
QY 64 DOEKFVVDENFTENYITDCEGKDAGNAGTGDESDEVD 102
Db 1063 D-----YLKDVNGTTSIGIGTSVEKYIIDE 1088

RESULT 12
US-09-930-871-10
; Sequence 10, Application US/09930871
; Patent No. US20020076780A1
; GENERAL INFORMATION:
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Mathur, Daniel
; APPLICANT: Mathur, Brian
; TITLE OF INVENTION: No. US20020076780A1el Human Ion Channel Proteins and Polynucle
; TITLE OF INVENTION: Same
; FILE REFERENCE: LEX-0216-USA
; CURRENT APPLICATION NUMBER: US/09/930,871
; PRIOR FILING DATE: 2001-08-14
; CURRENT APPLICATION NUMBER: US 60/225,989
; PRIOR FILING DATE: 2000-08-16
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 1387
; TYPE: PRT
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(1387)
; OTHER INFORMATION: xaa = Any Amino Acid
US-09-930-871-10

Query Match
Best Local Similarity 11.0%; Score 63; DB 10; Length 1387;
Matches 28; Conservative 10; Mismatches 31; Indels 30; Gaps 6;

QY 10 YANRKYTDF-DKSF-KKSSDDECKKTCFTECYIVFEDTVNKE--CYIVVD--GEEI 63
Db 1014 YVKKRIYEFIQOSFTRKQKILDEIKP-----LDDLNNKSCMSNHTXETGKDL 1062
QY 64 DOEKFVVDENFTENYITDCEGKDAGNAGTGDESDEVD 102
Db 1063 D-----YLKDVNGTTSIGIGTSVEKYIIDE 1088

RESULT 13
US-09-930-871-18
; Sequence 18, Application US/09930871
; Patent No. US20020076780A1
; GENERAL INFORMATION:
```

```

? APPLICANT: Turner, C. Alexander Jr.
? APPLICANT: Mathur, Daniel
? APPLICANT: Mathur, Brian
? TITLE OF INVENTION: No. US20020076780A1el Human Ion Channel Proteins and Polynucleot
? TITLE OF INVENTION: Same
? FILE REFERENCE: LEX-0216-USA
? CURRENT APPLICATION NUMBER: US/09/930, 871
? CURRENT FILING DATE: 2001-08-14
? PRIOR APPLICATION NUMBER: US 60/225,989
? PRIOR FILING DATE: 2000-08-16
? NUMBER OF SEQ ID NOS: 20
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 18
? LENGTH: 1392
? TYPE: PRT
? ORGANISM: homo sapiens
? FEATURE:
? NAME/KEY: VARIANT
? LOCATION: (1) --(1392)
? OTHER INFORMATION: xaa = Any Amino Acid
US-09-930-871-18

```

```

Query Match      11.0% ; Score 63; DB 10; Length 1392;
Best Local Similarity 28.3% ; Pred. No. 1.5e+02;
Matches 28; Conservative 10; Mismatches 31; Indels 30; Gaps 6;

QY    10 YANKRYTDF-DKSF-KRSSDLDECKKCTCFTEYCIVFEDVKNR--CYNVVD--GEEL 63
       | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db     1025 YVKRKRIEYIQSGSIFIKKQILDEIKP-----LDDLNKKKDCSMNHTYEIGKDL 1073

QY    64 DOEKFVVDENTENYLTDCEGKDAGNAAGTGSDSEVDE 102
       | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db     1074 D-----YLKDVNGTTSGIGCGSSSEKIIDE 1099
```

```

RESULT 14
US-09-930-871-20
: Sequence 20, Application US/09930871
Patent No. US20020076780A1
GENERAL INFORMATION:
APPLICANT: Turner, C. Alexander Jr.
APPLICANT: Machur, Daniel
APPLICANT: Machur, Brian
TITLE OF INVENTION: No. US20020076780A1el Human Ion Channel Proteins and Polynucleot
TITLE OF INVENTION: Same
FILE REFERENCE: LEX-0216-USA
CURRENT APPLICATION NUMBER: US/09/930, 871
CURRENT FILING DATE: 2001-08-14
PRIOR APPLICATION NUMBER: US 60/225,989
PRIOR FILING DATE: 2000-08-16
NUMBER OF SEQ ID NOS: 20
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 20
LENGTH: 1398
TYPE: PRT
ORGANISM: homo sapiens
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)-(1398)
OTHER INFORMATION: Xaa = Any Amino Acid
US-09-930-871-20

```

```

Query Match: 11.0%; Score 63; DB 10; length 1399;
Best Local Similarity 28.3%; Pred. NO. 1.5e+02;
Matches 28; Conservative 10; Mismatches 31; Indels 30; Gaps 6;

QY 10 YANKRYTF-DKSF-KKSDLDECKKTCRTEYCYVPEDEVNKE--CYNVVD--GEEL 63
      ||| :| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 1025 YVKRIIYEFQQGSFIKKKILDEIKP-----LDDLNNKDKSCMSNHYIEIGKDL 1073
      ||| :| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

QY 64 DQEKVVDENFTENTYLTDECGKADGAAAGTGSDEVDE 102
      ||| :| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 1074 D-----YLKDVNGTTSIGIGTSSVEKIITIDE 1099
      ||| :| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```

RESULT 15
 US-09-930-871-6
 : Sequence 6, Application US/09930871
 : Patent No. US20020076780A1
 : GENERAL INFORMATION:
 : APPLICANT: Turner, C. Alexander Jr.
 : APPLICANT: Mathur, Daniel
 : APPLICANT: Mathur, Brian
 : TITLE OF INVENTION: Same
 : FILE REFERENCE: LEX-0216-USA
 : CURRENT APPLICATION NUMBER: US/09/930,871
 : CURRENT FILING DATE: 2001-08-14
 : PRIOR APPLICATION NUMBER: US 60/225,989
 : PRIOR FILING DATE: 2000-08-16
 : NUMBER OF SEQ. ID NOS: 20
 : SOFTWARE: FastSeq for Windows Version 4.0
 : SEQ ID NO 6
 : LENGTH: 1442
 : TYPE: PRT
 : ORGANISM: homo sapiens
 : FEATURE:
 : NAME/KEY: VARIANT
 : LOCATION: (1)...(1442)
 : OTHER INFORMATION: Xaa = Any Amino Acid
 US-09-930-871-6

```

Query Match: 11.0%; Score 63; DB 10; Length 1442;
Best Local Similarity 28.3%; Pred. No. 1.6e+02;
Matches 28; Conservative 10; Mismatches 31; Indels 30; Gaps 6.

QY 10 YANRKYTDY-DKSF-KSSDLDCEKCKTCFKTEYCYIVFEDTVNKE--CYNVVD--GEET 63
      | | : : | | : | | | : | | : | | : | | : | | : |
Db 1014 YVKRKIIYEFQGSFIRKQILDEIRP-----LDINNKKDCSMNHYEIGKDL 1062

QY 64 DQEKVVDENFTENYLTDCGKADAGNAAGTGDESDVE 102
      | | : | : | : | : | : | : | : | : | : | : |
Db 1063 D-----YKDVNGTTSIGIGGSSVEKIIIDE 1088

```

Search completed: February 25, 2003, 02:34:14
Job time : 13 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 25, 2003, 02:32:55 ; Search time 2615 Seconds
(without alignments)
1146.306 Million cell updates/sec

Title: US-09-936-737A-2
Perfect score: 573
Sequence: 1 EEREDCWTYANKRYTDFDK.....GMDAGNAGTDESDDEVDED 103

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 segs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame_plus_p2n.model -DEV=xlh
-Q=CGN2_1/USPTO_SPOOL/US0996737/runat.14022003.100732.2291/app.query.fasta.1.263
-DB=GenEmbl -QFWT=FASTA -SUFFIX=rig -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=PCT -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFWT=PRO -NORM=EXT -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0996737.CGN.1.1.2425.0runat.14022003.100732.2291 -NCPD=6 -ICPU=3
-NO_XIPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7
-Ygapop=10 -Ygapext=0.5 -Delop=6 -Delext=7

Database :

GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_in:*
18: em_hum:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	573	100.0	375	6 AX036541	AX036541 Sequence
2	90	15.7	110000	2 PFMAL13P2_1	Continuation (2 of
3	90	15.7	110000	2 PFMAL13P2_2	Continuation (3 of
4	88	15.4	138024	2 AC131381	AC131381 Strongylo
5	86.5	15.1	128960	9 HSJ919F19	AL109947 Human DNA
6	85.5	14.9	563	9 HEITAPP	M81489 Haemeteria
7	83	14.5	95615	9 AL138835	AL138835 Human DNA
8	83	14.5	149109	2 AC101018	AC101018 Rattus no
9	83	14.5	159492	2 AC115173	AC115173 Rattus no
10	81.5	14.2	2255	6 SFU15038	U15038 Spodoptera
11	81.5	14.2	2255	6 AR030964	AR030964 Sequence
12	80.5	14.0	178681	2 PFDMACP	X84041 P.falciparu
13	80.5	13.9	110000	2 AC098520	AC098520 Rattus no
14	79.5	13.9	110000	2 AC108272_2	Continuation (3 of
15	79.5	13.9	232370	2 AC103570	AC103570 Rattus no
16	79	13.8	43380	3 CER2604A	AL110478 Caenorhab
17	78.5	13.7	7311	1 SPN131985	AT131985 streptococ
18	78.5	13.7	32110	3 CER40612	Z77661 Caenorhabdi
19	78.5	13.7	212085	10 AL663030	AL663030 Mouse DNA
20	78.5	13.7	301550	1 AP003192	AP003192 Clostridi
21	78	13.6	44237	3 AC024847	X64346 Herpesvirus
22	78	13.6	112330	14 HSGEND	AL034558 Plasmodiu
23	78	13.6	158548	3 PFMAL13P2	AL356116 Homo sapi
24	78	13.6	174280	2 AL356116	AC006889 Caenorhab
25	78	13.6	267118	2 AC006889	AP001517 Bacillus
26	78	13.6	294250	1 AP001517	AL049184 Plasmodiu
27	78	13.6	318221	2 PFMAL13P3	AB019231 Arabidops
28	77.5	13.5	43570	8 AB019231	AL445568 Human DNA
29	77.5	13.5	46341	2 AL445568	AC006281 Plasmodiu
30	77.5	13.5	199531	2 AC006281	M55686 T.brucei ac
31	77	13.4	500	3 TRECARM1	AY071572 Drosophil
32	77	13.4	2625	3 AY071572	AP004382 Oryza sat
33	77	13.4	141040	8 AP004382	U58762 Caenorhabdi
34	76.5	13.4	34960	3 U58762	U20549 Restreptia m
35	76	13.3	2252	8 RMU20549	AC084663 Caenorhab
36	76	13.3	35766	3 CBMG47C05	Z68135 Caenorhabdi
37	76	13.3	37640	3 CER27C14	Z68135 Caenorhabdi
38	76	13.3	121290	2 AC111055	AC111055 Mus muscu
39	76	13.3	159001	2 AC012139	AC068672 Mus muscu
40	76	13.3	164748	9 AC068672	AC007305 Mus muscu
41	76	13.3	195883	2 AC007305	AR102119 Sequence
42	76	13.3	214839	6 AR102119	AR103163 Sequence
43	75.5	13.2	1402	6 AR103163	BC028979 Mus muscu
44	75.5	13.2	1402	6 AR103163	
45	75.5	13.2	2228	10 BC028979	

RESULT 1

ALIGNMENTS

BASE COUNT	123 a	62 c	85 g	105 t
ORIGIN				

Alignment Scores:	
Pred. No.:	4,21e-60
Score:	573.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	100.00%
DB:	6
Length:	375
Matches:	103
Conservative:	0
Mismatches:	0
Indels:	0
Gaps:	0

US-09-936-737A-2 (1-103) x AX036541 (1-375)

QY	1	GIUGIUGAGCUGAUSPCYSTRPHRPhenylalalasnaryslyrThraspPheaspLys	20
Db	64	GAAAGACGTGAACATTCTTTGACGTTTACCGGAACGAAAAATATACACTTCATTA	122
QY	21	SerPheLysLysSerSerAspLeuaspLUCysLysLysThrCysPheLysThrLunyr	40
Db	124	TCCTTTTAAGAACGCCCTCATCTTGACGAATGCAGAAAAACAAGTTTCAACGCGAGTAC	18
QY	41	CysTyrLLeValPheGluaspThrValasnLysGluCysTyrTyrAsnValaspLys	60
Db	184	TGCTACACTCCTTTTGTGACACACGCTCAACAGGAATGTTACTACAAATGCTGTGATGCT	243
QY	61	GluGluLeuaspGluGluLysPheValaspLysAsnPheThrGluAsnTyrLeuThr	80
Db	244	GAAAGAGTGAACCAAGAAAAATTTGTTGTCGACAGAAACTTCACGAAAAATATTATTACA	303
QY	81	AspCysGluGluLysAspAlaGluAsnAlaIaIaGlyThmGlyAspGluSerAspGluVal	100
Db	304	GACTCGAGGAGTAAGAATCGAGCTATATCGCGCAGGTACAGGTGACGACTCGAGTGAAGATT	363
QY	101	AspAspLasp 103	
Db	364	GATGAAGAT 372	

RESULT 2				
PFMAL13p2_1/c				
WPCOMMENT				
Sequence split into 4 fragments				
Fragment Name	Begin	End		
PFMAL13p2_0	1	110000		
				AL049185

Alignment Scores:
Pred No.:

US-09-936-737A-2 (1-103) x PFMAL13P2_1 (1-110000)

.....GAGATTTATTAATAATAATAA 101438

Pb 101437 TTTTACATTTTTCTATTGAAAGGCGGGGTTC
|||||||:::||
LEUASpglu 30 ::||

[illegible][illegible]

QY 48 ThrValAsnLysGluCysTyrTyrAsnValVal----- 58

Db 101317 GTATTAGGTAAGTCTGATTATATTAATCAATGGAGATATAAAATTAAGAAGACT 101258

58

CAACATGATTTGATTAATTTAGATTTTGATTAATCATGATGTGAGCAA 101198

b 101107 :cmmccccc-----
-aspelyglugluLeuaspcnglLyspheval 69
::: :::::||| :::
::: :::::||| :::

70 ValAsnGluAspSerHisProValLeuTyrIle

-----CNCNCGAGATTCATACTATT 101138

db 101137 GTGATAAAATATATCAGACAATCTATT
:::
:::
:::

90 AlaAlaGlyThrGlyAspGlnSerAspGlnValAsnGln 102

bp 101092 GAAGCTACAGAAAAGACGAAAAGATCAAGTGTATGAA 101054

RESULT 3
FMAL13P3 2/5

COMMENT:
Sequence split into 4 fragments 100ms duration

Logarithmic Name	Begin	End
PFMAL13P2_0	1	110000

1000001	210000
2000001	310000

Continuation (3 of 4) of PFMAL13P2 from base 200001 (AL049185 Plasmodium falci-

Alignment Scores: 33 5

Percent Similarity:	90.00	Matches:	33
	34.648	Conservation:	20

Identity:	21.3%	Mismatches:	40
Query Match:	15.71%	Indels:	60

9
odps:

5 AsnGvstTrrbhbphctuuat

1497 GATATATGTAACCTTCATAAAAGTCATAATTATCCTTCTTTAATT

18 -----PheaspLysSerPheLysLysSerSerAsp-----Iouhacj 30

* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.

```

* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
* 1 58994: contig of 38994 bp in length
* 58995 59094: gap of unknown length
* 59095 96228: contig of 37134 bp in length
* 96229 96328: gap of unknown length
* 96329 138024: contig of 41696 bp in length.
FEATURES
    source
        1..138024
            /organism="Strongylocentrotus purpuratus"
            /db_xref="taxon:7668"
            /clone="Sp31u8"
            /clone.lib="Caltech Strongylocentrotus purpuratus sperm
            genomic BAC library A"
            /note="This library is described in Cameron,R.A.,
            Mahairas,G., Rast,J.P., Martinez,P., Blondi,T.R.,
            Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,
            ,B.T., Wray,G.A., Etlenson,C.A., Lehrach,H., Britten,R.J
            Davidson,E.H., Hood,L. Proc. Natl. Acad. Sci. U.S.A. 97
            (17): 9514-8 2000"
BASE COUNT  42761 a 25096 c 25960 g 43507 t 700 others
ORIGIN
Alignment Scores:
Pred. No.:      76.2      Length:      138024
Score:          88.00     Matches:      18
Best Local Similarity: 52.83%  Conservative: 10
Percent Local Similarity: 33.96%  Mismatches:   7
Query Match:      15.36%      Indels:      18
                        Gaps:      2
DB:
US-09-936-737A-2 (1-103) x AC131381 (1-138024)
QY
QY 31 CysLysLysThr-----CysPheLysrHrGlu 39
      ||| :|||
      TGCAGCGCGACAGCACTCTAGGACAAAGCATGCTGTCAGTGCCTAAACATCAAA 133127
DB 133186
QY 40 TyrCysTyrIleValPheGluAspThrValAsnLysGluCysTyrTyrAsnValVal--- 58
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      TACGTGTTCAATTTATTTTAAACACGCTCTAAATAAAGAAATGTAATTAATCTCATAT 133067
DB 133126
QY 59 -----AspGlyLugLugLueLusAspGln 65
      ::|||::: |||||:::
      GACCAACATCTGCACCAATGGGAAGATCTTGATAG 133028
DB 133066
RESULT 5
HSJ919F19/C
LOCUS      HSJ919F19
DEFINITION Human DNA sequence from clone Rp5-919F19 on chromosome 6q16.3-22.1,
complete sequence.
ACCESSION AL109947
VERSION    AL109947.19  GI:13897493
KEYWORDS   human.
SOURCE     Homo sapiens
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE  1 (bases 1 to 128960)
AUTHORS   Johnson,C.
TITLE     Direct Submission
JOURNAL   Submitted (11-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clone.requests@sanger.ac.uk
On Apr 30, 2001 this sequence version replaced gi:13751877.
COMMENT

```

chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL, Sw: SWISSPROT, Tr: TREMBL, Mp: MORPEP; Information on the MORPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep. This sequence was generated from part of bacterial clone contigs of human Chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr6>. RP5-919P19 is from the library RPCI-5 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>. VECTOR: pCYPAC2

This sequence is the entire insert of clone RP5-919P19. The true left end of clone RP5-1036D8 is at 86977 in this sequence. The true right end of clone RP11-425D10 is at 56374 in this sequence.

FEATURES

Source

Location/Qualifiers
1..128960
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="q16.3-22.1"
/clone="RP5-919P19"
/clone_11p="RPCI-5"
16..330
repeat_region
/note="AluY repeat: matches 1..298 of consensus"
878..1177
repeat_region
/note="AluY repeat: matches 1..301 of consensus"
1409..1704
repeat_region
/note="AluX repeat: matches 1..296 of consensus"
2306..2424
repeat_region
/note="FLAM_A repeat: matches 1..123 of consensus"
3935..4225
repeat_region
/note="AluSg repeat: matches 1..299 of consensus"
4512..4698
repeat_region
/note="L1MC5 repeat: matches 7347..7537 of consensus"
5185..5506
repeat_region
/note="L1MC5 repeat: matches 1..308 of consensus"
5558..6086
repeat_region
/note="AluDo repeat: matches 1..308 of consensus"
6137..6330
repeat_region
/note="L1MC5 repeat: matches 6834..7364 of consensus"
6137..6330
repeat_region
/note="L1MC/D repeat: matches 5276..5467 of consensus"
6393..6686
repeat_region
/note="AluY repeat: matches 1..301 of consensus"
6998..7112
repeat_region
/note="FLAM_A repeat: matches 1..115 of consensus"
7186..7588
repeat_region
/note="HA1 repeat: matches 505..914 of consensus"
7998..8355
repeat_region
/note="L1MA2 repeat: matches 3613..3976 of consensus"
8372..8434
repeat_region
/note="L1MA2 repeat: matches 4133..4193 of consensus"
8435..8737
repeat_region
/note="AluX repeat: matches 1..298 of consensus"
8738..9144
repeat_region
/note="L1MA2 repeat: matches 4193..4614 of consensus"
9339..11100
repeat_region
/note="MER66-internal repeat: matches 4895..6676 of consensus"
11101..11496
repeat_region
/note="MER66B repeat: matches 1..392 of consensus"
11510..12032
repeat_region
/note="L1MA2 repeat: matches 4604..5114 of consensus"
12033..12338
repeat_region
/note="AluSg repeat: matches 1..306 of consensus"
12339..13410
repeat_region
/note="L1MA2 repeat: matches 5114..6307 of consensus"
13633..13852
repeat_region

repeat_region
/note="AluDb repeat: matches 87..307 of consensus"
13896..15355
repeat_region
/note="TRIGER2 repeat: matches 1266..2718 of consensus"
15356..15657
repeat_region
/note="AluX repeat: matches 1..301 of consensus"
15658..16040
repeat_region
/note="TRIGER2 repeat: matches 863..1266 of consensus"
16041..16350
repeat_region
/note="AluDo repeat: matches 1..306 of consensus"
16351..16417
repeat_region
/note="TRIGER2 repeat: matches 798..863 of consensus"
16445..16538
repeat_region
/note="FLAM_C repeat: matches 24..117 of consensus"
16579..16625
repeat_region
/note="TRIGER2 repeat: matches 1..47 of consensus"
16749..17174
repeat_region
/note="MER65A repeat: matches 1..445 of consensus"
17696..17996
repeat_region
/note="AluX repeat: matches 1..306 of consensus"
18497..18587
repeat_region
/note="FLAM_A repeat: matches 43..130 of consensus"
18596..18837
repeat_region
/note="AluSg repeat: matches 1..302 of consensus"
19209..19517
repeat_region
/note="AluX repeat: matches 1..312 of consensus"
20539..20659
repeat_region
/note="MIR repeat: matches 98..228 of consensus"
20912..21038
repeat_region
/note="AluDo repeat: matches 1..126 of consensus"
21039..21330
repeat_region
/note="AluSg repeat: matches 1..291 of consensus"
21331..21467
repeat_region
/note="AluDo repeat: matches 126..266 of consensus"
21473..21756
repeat_region
/note="AluDo repeat: matches 14..295 of consensus"
22240..22524
repeat_region
/note="AluX repeat: matches 1..286 of consensus"
23046..23585
repeat_region
/note="L12 repeat: matches 2234..2745 of consensus"
23679..24214
repeat_region
/note="MER41A repeat: matches 1..538 of consensus"
24274..24440
repeat_region
/note="MIR repeat: matches 17..191 of consensus"
24622..24927
repeat_region
/note="AluX repeat: matches 1..305 of consensus"
26536..26843
repeat_region
/note="AluDb repeat: matches 1..303 of consensus"
27436..27726
repeat_region
/note="AluDb repeat: matches 13..304 of consensus"
29235..29339
repeat_region
/note="L12 repeat: matches 2596..2705 of consensus"
29588..29879
repeat_region
/note="AluSc repeat: matches 1..293 of consensus"
30927..31173
repeat_region
/note="L1R23 repeat: matches 204..437 of consensus"
31638..31785
repeat_region
/note="MER5A repeat: matches 87..189 of consensus"
31786..31914
repeat_region
/note="FLAM_C repeat: matches 1..131 of consensus"
31915..31969
repeat_region
/note="MER5A repeat: matches 2..87 of consensus"
32211..32511
repeat_region
/note="AluX repeat: matches 1..301 of consensus"
33285..33395
repeat_region
/note="MER81 repeat: matches 3..114 of consensus"
33429..33566
repeat_region
/note="FLAM_C repeat: matches 1..138 of consensus"
33567..33606
repeat_region
/note="20 copies 2 mer tg 87% conserved"
34452..34750
repeat_region
/note="AluSc repeat: matches 1..299 of consensus"
34798..35089
repeat_region
/note="AluDb repeat: matches 1..292 of consensus"

```

repeat_region      35591..35905
                    /note="AluX repeat: matches 1..312 of consensus"
repeat_region      35917..35978
                    /note="L2 repeat: matches 2622..2650 of consensus"
repeat_region      36099..37024
                    /note="AluX repeat: matches 1..311 of consensus"
repeat_region      37215..37349
                    /note="L1M4S4 repeat: matches 6146..6295 of consensus"
repeat_region      37451..37586
                    /note="L1M4S4 repeat: matches 6146..6295 of consensus"
repeat_region      37718..37742
                    /note="L1M4S4 repeat: matches 6146..6295 of consensus"
repeat_region      37743..38041
                    /note="L1M4S4 repeat: matches 6146..6295 of consensus"
repeat_region      38042..38272
                    /note="L1M4S4 repeat: matches 6146..6295 of consensus"
repeat_region      38504..38786
                    /note="L1M4S4 repeat: matches 6146..6295 of consensus"
misc_feature        39348..40556
                    /note="CpG island"
misc_feature        /evidence=not_experimental
                    /complement(42531..43073)
misc_feature        /note="match: STS: Em:636500"
repeat_region      43639..45763
                    /note="L2 repeat: matches 2617..2750 of consensus"
repeat_region      47289..47314
                    /note="L3 repeat: matches 2617..2750 of consensus"
repeat_region      47413..47506
                    /note="L1M2 repeat: matches 128..244 of consensus"

```

Alignment Scores:

Pred. No.:	107	Length:	128960
Score:	86.50	Matches:	32
Percent Similarity:	42.11%	Conservative:	16
Best Local Similarity:	28.07%	Mismatches:	49
Query Match:	15.10%	Indels:	17
		Gaps:	5

US-09-936-737a-2 (1-103) x HST919F19 (1-128960)

```

QY 3 ArgGluSpCysTrpThrPheTyrAlaAsnArgLysTyrThrAspPheAspLysSerPhe 22
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 115992 CACGCTGACACAAAAGTCTTAAACCCCAAGTGTGAGATATTAAACCTTAT 115933
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 23 LysLysSerSerAspLeuAspGluCysLysLysThrCysPheLys-----ThrGluTyr 40
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 115932 TGGGTTACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 115973
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 41 -----CysTyrLleValPheGluAspThrValAsnLys 51
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 115872 TCCAGATGATTTGTTATAGTGTCTTTGTTACATATTAT-----ATGTTATATAA 115819
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 52 GluCysTyrTyr-----AsnValVal-----AspGluGluGluLeuAspGluGluLys 67
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 115818 AGGCAATATTTTCTTCCCTAGATGATGATGATGATGATGATGATGATGATGATGAT 115759
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 68 PheValValAspGluAspLeuThrGluAsnTyrLeuThrAspCysGluGluLysAspAla 87
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 115758 CTGAGAGACACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 115699
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 88 GlyAsnAlaAlaGlyThrGlyAspGluSerAspGluValAsp 101
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 115698 GAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 115657
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

RESULT 6

HEILAPP 563 bp mRNA linear INV 26-APR-1993
 LOCUS Haementeria officialis anti-platelet protein (LAPP) mRNA, complete cds.
 ACCESSION M81489
 VERSION M81489.1 GI:159220
 KEYWORDS anti-platelet protein.
 SOURCE Haementeria officialis cDNA to mRNA.

ORGANISM Haementeria officialis
 Eukaryota; Metazoa; Annelida; Clitellata; Hirudinida; Hirudinea;
 Rhynchobdellida; Glossiphoniidae; Haementeria.
 REFERENCE 1 (bases 1 to 563)
 AUTHORS Keller, P.M., Schultz, L.D., Condra, C., Karzewski, J. and Connolly, T.M.
 TITLE An inhibitor of collagen-stimulated platelet activation from the salivary glands of the Haementeria officialis leech. II. Cloning of the cDNA and expression
 JOURNAL J. Biol. Chem. 267 (10), 6899-6904 (1992)
 MEDLINE 92202247
 PUBMED 1551898
 FEATURES
 source location/Qualifiers
 1..563 /organism="Haementeria officialis"
 /db_xref="taxon:6410"
 gene 1..563 /gene="LAPP"
 CDS 58..501 /gene="LAPP"
 /codon_start=1
 /product="anti-platelet protein"
 /protein_id="AA29194.1"
 /db_xref="GI:159221"
 /translation="MNSPLPSLACSLVLAIPAISNAPEDAGADETSEGETTGSPE
 TPSTGGGDCGNETITRGNDCSKRPMGLPUNLITKEFTSVDECRKCKESAVE
 PSCYLLQINTETNECYRNEDGVWSSLDQPNVGMHLHACSK"
 58..120 /gene="LAPP"
 121..498 /gene="LAPP"
 /product="anti-platelet protein"
 /gene="LAPP"
 563 /product="anti-platelet protein"
 polyA-site
 BASE COUNT 190 a 114 c 127 g 132 t
 ORIGIN

Alignment Scores:

Pred. No.:	0.24	Length:	563
Score:	85.50 <td>Matches:</td> <td>19 </td>	Matches:	19
Percent Similarity:	51.79% <td>Conservative:</td> <td>10 </td>	Conservative:	10
Best Local Similarity:	33.93% <td>Mismatches:</td> <td>24 </td>	Mismatches:	24
Query Match:	14.92% <td>Indels:</td> <td>3 </td>	Indels:	3
		Gaps:	1

US-09-936-737a-2 (1-103) x HEILAPP (1-563)

```

QY 4 GluAspCysTrpThrPheTyrAlaAsnArgLysTyrThrAspPheAspLysSerPheLys 23
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 250 GAAGATGCTGCTGCTAAACGCGCTGCGGAGAGCTGCTGACATCTGTTGACGAGACT 309
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 24 LysSerSerAspLeuAspGluCysLysLysThrCysPheLysThrGluTyr----- 40
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 310 GAATTCACGAGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 369
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 41 CysTyrLleValPheGluAspThrValAsnLysGluCysTyrTyrAsn 56
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 370 TECTACATTCGCAATGACACAGAAACCAAGCATCTATAGAAAT 417
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

RESULT 7

AL138835 95615 bp DNA linear PRI 13-APR-2001
 LOCUS Human DNA sequence from clone RP11-554F20 on chromosome 9, complete sequence.
 ACCESSION AL138835
 VERSION AL138835.9 GI:13624930
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 95615)
 AUTHORS Lloyd, D.

TITLE Direct Submission
JOURNAL Submitted (13-APR-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
COMMENT requests: clonerequest@sanger.ac.uk
On Apr 14, 2001 this sequence version replaced g1:13442296.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WormPeP; Information on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr9>
Rp11-554P20 is from the library RPCT-11.2 constructed by the group of Pieter de Jong. For further details see
<http://www.choil.org/bacpac/home.htm>
VECTOR: pBAC3.6
IMPORTANT: This sequence is not the entire insert of clone Rp11-554P20. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.
The true left end of clone Rp11-554P20 is at 1 in this sequence. The true left end of clone Rp11-30L7 is at 9516 in this sequence. The true right end of clone Rp11-547C13 is at 37658 in this sequence.

FEATURES

source

1..95615 location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="Rp11-554P20"
/clone_lib="RPCT-11.2"
114..351
/note="12 repeat: matches 2341..2566 of consensus"
387..846
/note="M11C repeat: matches 21..466 of consensus"
1111..1407
/note="Alu10 repeat: matches 7..308 of consensus"
1864..2320
/note="match: STS: Em:HSJ61D7"
1881..1925
/note="5 copies 9 mer tgtctggtg 88% conserved"
2094..2286
/note="12 repeat: matches 2009..2207 of consensus"
3322..3470
/note="MIR repeat: matches 61..207 of consensus"
3943..4199
/note="12 repeat: matches 2423..2702 of consensus"
4307..4480
/note="MER58A repeat: matches 37..207 of consensus"
4517..4617
/note="12 repeat: matches 123..232 of consensus"
5850..5932
/note="MIR repeat: matches 1988..2234 of consensus"
6248..6480
/note="MER96 repeat: matches 91..174 of consensus"
7258..7554
/note="12 repeat: matches 1988..2234 of consensus"
8289..8499
/note="Alu10 repeat: matches 1..297 of consensus"
/note="MIR repeat: matches 8..237 of consensus"

repeat_region 8819..8901
/note="12 repeat: matches 2335..2419 of consensus"
repeat_region 8927..8996
/note="12 repeat: matches 2231..2299 of consensus"
repeat_region 9012..9531
/note="12 repeat: matches 2180..2738 of consensus"
repeat_region 9586..9894
/note="12 repeat: matches 2348..2747 of consensus"
repeat_region 10360..10424
/note="12 repeat: matches 2646..2708 of consensus"
repeat_region 10421..10489
/note="MIR repeat: matches 48..118 of consensus"
repeat_region 11504..11584
/note="MIR repeat: matches 184..260 of consensus"
repeat_region 12099..12256
/note="MIR repeat: matches 16..190 of consensus"
repeat_region 13382..13680
/note="MIR repeat: matches 143..457 of consensus"
repeat_region 13681..14104
/note="MIR repeat: matches 65..527 of consensus"
repeat_region 14129..14174
/note="MIR repeat: matches 411..495 of consensus"
repeat_region 15322..15616
/note="12 repeat: matches 2656..2749 of consensus"
repeat_region 15613..15829
/note="12 repeat: matches 2177..2418 of consensus"
repeat_region 17533..17662
/note="12 repeat: matches 2613..2750 of consensus"
repeat_region 18242..18593
/note="MIR repeat: matches 1..365 of consensus"
repeat_region 19072..19266
/note="MIR repeat: matches 118..319 of consensus"
repeat_region 19843..20184
/note="MIR repeat: matches 65..410 of consensus"
repeat_region 20734..20959
/note="MIR repeat: matches 10..250 of consensus"
repeat_region 21671..22060
/note="MIR repeat: matches 5777..6165 of consensus"
repeat_region 22327..22470
/note="MIR repeat: matches 13..163 of consensus"
repeat_region 22860..22957
/note="MIR repeat: matches 47..146 of consensus"
repeat_region 23642..23705
/note="MIR repeat: matches 68..132 of consensus"
repeat_region 24151..24345
/note="MIR repeat: matches 29..235 of consensus"
repeat_region 24365..24964
/note="MIR repeat: matches 5660..6286 of consensus"
repeat_region 24986..25254
/note="12 repeat: matches 2459..2747 of consensus"
repeat_region 25934..26141
/note="MIR repeat: matches 1..253 of consensus"
repeat_region 26508..26640
/note="MIR repeat: matches 124..245 of consensus"
repeat_region 26655..26892
/note="MIR repeat: matches 6..262 of consensus"
repeat_region 26900..26992
/note="MIR repeat: matches 20..111 of consensus"
repeat_region 27024..27480
/note="12 repeat: matches 1815..2289 of consensus"
repeat_region 27873..28179
/note="Alu10 repeat: matches 1..305 of consensus"
repeat_region 28301..28417
/note="MIR repeat: matches 1..122 of consensus"
repeat_region 28394..29077
/note="MIR repeat: matches 51..112 of consensus"
repeat_region 29079..29136
/note="29 copies 2 mer at 70% conserved"
repeat_region 29188..29307
/note="12 repeat: matches 2625..2748 of consensus"
repeat_region 29427..29487
/note="MIR repeat: matches 51..112 of consensus"

```

repeat_region 29591..29906
                /note="AluY repeat: matches 1..310 of consensus"
repeat_region 30323..30712
                /note="MUTB repeat: matches 7..390 of consensus"
repeat_region 33200..33272
                /note="L2 repeat: matches 2626..2702 of consensus"
repeat_region 34778..34890
                /note="MER5A repeat: matches 3..110 of consensus"
repeat_region 35583..35685
                /note="MIR repeat: matches 81..180 of consensus"
repeat_region 37170..37243
                /note="MIR repeat: matches 77..148 of consensus"
repeat_region 37581..38098
                /note="L2 repeat: matches 2153..2750 of consensus"
repeat_region 38695..38938
                /note="MIR repeat: matches 21..261 of consensus"
repeat_region 38938..39002
                /note="MER5A repeat: matches 97..161 of consensus"
repeat_region 39559..39602
                /note="L2 repeat: matches 2638..2720 of consensus"
repeat_region 40801..40832
                /note="L6 copies 2 mer ga 84% conserved"
repeat_region 40985..41325
                /note="AluYB8 repeat: matches 1..318 of consensus"
repeat_region 41844..42158
                /note="AluX repeat: matches 1..310 of consensus"
repeat_region 42382..42535
                /note="MIR repeat: matches 95..252 of consensus"
repeat_region 42788..42921
                /note="L2 repeat: matches 2555..2691 of consensus"
repeat_region 43117..43172
                /note="L2 repeat: matches 2555..2691 of consensus"
repeat_region 43604..43697
                /note="L28 copies 2 mer tt 73% conserved"
repeat_region 43698..43818
                /note="L1M9 repeat: matches 5693..5786 of consensus"
repeat_region 43860..44067
                /note="Tigfz (Zombi) repeat: matches 3..124 of consensus"
repeat_region 44088..44127
                /note="MST repeat: matches 1..229 of consensus"
repeat_region 44129..44338
                /note="20 copies 2 mer tc 90% conserved"
repeat_region 44129..44338
                /note="MST repeat: matches 200..426 of consensus"

```

```

Alignment Scores:
Pred. No.: 197
Score: 83.00
Percent Similarity: 48.61%
Best Local Similarity: 36.11%
Query Match: 14.49%
DB: 9
Gaps: 2

```

US-09-936-737a-2 (1-103) x AL138835 (1-95615)

```

QY 4 GluaspCysTrpThrPheTyAlaAsnArgLysThrAspPheAspLysSerPheLys 23
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 95304 GAGCAGCTAGGAGGCTAGAGGCGCCAGCACTCTATATGATTTTAAAAAGCA----- 95357
QY 24 LysSerSerAspLysAspGluCysLysLysThrCysPheLysThrGluLys----- 40
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 95358 ---GCCCTCAGAGATGCTGAGAGTTTGGATTCGTTGTAAGAAAGAGGTTAAGA 95414
QY 41 CysTrpLysValPheGluAspThrValAsnLysGluCysTrpLysValValAspGly 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 95415 TGTACACAAATTTTGGCCCTGGAACAGCAAGAAAGAGATTCATGCTGCTGAAGG 95474
QY 61 GluGluLeuAspGluLysPheValValAspGlu 72
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 95475 GAGCACTGTAAAGAGAGTGTGATGAGCA 95510

```

RESULT 8 AC101018 149109 bp DNA linear HTG 21-AUG-2002
 LOCUS AC101018
 DEFINITION Rattus norvegicus clone RP32-328P7, *** SEQUENCING IN PROGRESS ***,

ACCESSION VERSION KEYWORDS SOURCE ORGANISM

REFERENCE AUTHORS TITLE JOURNAL REFERENCE AUTHORS

TITLE JOURNAL REFERENCE AUTHORS

TITLE JOURNAL COMMENT

9 unordered pieces.
 AC101018
 AC101018.2 GI:22381320
 HTG: HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEPIN.
 Norway rat.
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 149109)
 Birren,B., Nusbaum,C. and Lander,E.
 Rattus norvegicus, clone RP32-328P7
 Unpublished
 2 (bases 1 to 149109)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., All,A., Allen,N.,
 Anderson,S., Bana,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
 Brown,A., Camarata,U., Campolano,A., Chang,J., Chazaro,B.,
 Choepel,Y., Collangelo,M., Collins,S., Collymore,A., Cook,A.,
 Cooke,P., Dearlano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
 Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
 Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
 Hages,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
 Jones,C., Kamat,A., Karatas,A., Kells,C., Lacroque,K.,
 Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
 Maclean,C., Macdonald,P., Major,J., Margulis,N., Matthews,C.,
 McCarthy,M., McEwan,P., McKernan,K., McPheters,R., Meldrim,J.,
 Menus,L., Mihova,T., Mlenka,V., Murphy,T., Naylor,J., Nguyen,C.,
 Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
 Oliver,J., Peterson,K., Phunhphang,P., Pierre,N., Pollara,V.,
 Raymond,C., Retta,R., Riedack,M., Riley,R., Rise,C., Rogov,P.,
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R.,
 Seaman,S., Severy,J., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Strauss,N., Subramanian,A., Talmas,J., Tesfaye,S., Theodore,J.,
 Topham,K., Travers,M., Travis,N., Trigglio,J., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 149109)
 Birren,B., Nusbaum,C., Lander,E., All,A., Allen,N., Anderson,S.,
 Bana,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B.,
 Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
 Cook,A., Cooke,P., Dearlano,K., Dewar,K., Diaz,J.S., Dodge,S.,
 Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J.,
 Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hages,B.,
 Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
 Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
 Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
 McCarthy,M., Meldrim,J., Menus,L., Mihova,T., Mlenka,V.,
 Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
 O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
 Phunhphang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
 Roman,J., Roy,A., Schauer,S., Schuback,R., Seaman,S., Severy,P.,
 Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talmas,J.,
 Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
 Zembek,L., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 All Aug 21, 2002 this sequence version replaced g1117059792.
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RW/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIMB
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L14794
 Center clone name: 328_P7

*	1	1014:	contig of 1014 bp in length
*	1015	1114:	gap of unknown length
*	1115	2395:	contig of 1281 bp in length
*	2396	2495:	gap of unknown length
*	2496	3600:	contig of 1105 bp in length
*	3601	3700:	gap of unknown length
*	3701	5021:	contig of 1321 bp in length
*	5022	5121:	gap of unknown length
*	5122	6727:	contig of 1606 bp in length
*	6728	6627:	gap of unknown length
*	6628	8608:	contig of 1781 bp in length
*	8609	8708:	gap of unknown length
*	10004	10003:	contig of 1295 bp in length
*	10104	11013:	gap of unknown length
*	11121	11211:	contig of 1108 bp in length
*	11312	12518:	contig of 1207 bp in length
*	12519	12618:	gap of unknown length
*	12619	13843:	contig of 1225 bp in length
*	13844	13943:	gap of unknown length
*	13944	15404:	contig of 1461 bp in length
*	15405	15504:	gap of unknown length
*	15505	17647:	contig of 2143 bp in length
*	17648	17747:	gap of unknown length
*	17748	19706:	contig of 1959 bp in length
*	19707	19806:	gap of unknown length
*	19807	21576:	contig of 1770 bp in length
*	21577	21576:	gap of unknown length
*	21677	22906:	contig of 1230 bp in length
*	22907	23006:	gap of unknown length
*	23007	25086:	contig of 2080 bp in length
*	25087	25186:	gap of unknown length
*	25187	26832:	contig of 1646 bp in length
*	26833	26932:	contig of 1778 bp in length
*	26933	26932:	gap of unknown length
*	28711	28710:	contig of 1778 bp in length
*	28811	28810:	gap of unknown length
*	30606	30606:	contig of 1796 bp in length
*	30707	30706:	gap of unknown length
*	33113	33113:	contig of 2407 bp in length
*	33114	33213:	gap of unknown length
*	33214	35361:	contig of 2148 bp in length
*	35362	35461:	gap of unknown length
*	35462	37520:	contig of 2059 bp in length
*	37521	37520:	gap of unknown length
*	37621	40471:	contig of 2851 bp in length
*	40472	40571:	gap of unknown length
*	40572	42584:	contig of 2013 bp in length
*	42585	42684:	gap of unknown length
*	42685	45547:	contig of 2863 bp in length
*	45548	45647:	gap of unknown length
*	45648	47838:	contig of 2191 bp in length
*	47939	47938:	gap of unknown length
*	47939	51229:	contig of 3291 bp in length
*	51230	51329:	gap of unknown length
*	51330	54005:	contig of 2676 bp in length
*	54006	54105:	gap of unknown length
*	54106	56683:	contig of 2578 bp in length
*	56684	56783:	gap of unknown length
*	56784	59361:	contig of 2578 bp in length
*	59362	59461:	gap of unknown length
*	59462	61588:	contig of 2127 bp in length
*	61589	61688:	gap of unknown length
*	61689	65712:	contig of 4024 bp in length

Source	Location/Qualifiers	Base Count	Origin
1. .159492	/db_xref="taxon:10116"	39083	a
/organism="Rattus norvegicus"	/clone="CH230-286017"	36447	c
		36055	g
		40464	t
		7443	other

Pred. No.:	360	Length:	83.00
Score:	83.00	Matches:	Conservative:
Percent Similarity:	50.00%		Mismatches:
Best Local Similarity:	30.30%		Indels:
Query Match:	14.49%		Gaps:
DB:	2		

Db	110192	GACTCACTGTTATGTCAGCAGGAGAAGAGGTGGAGCGACGACTCTCATGCAGATTAC	11
QY	59	AspglyGIUGLIGLeuaSpGLInGUySPheValValaspGLuan---PheThrGLuan	77
Db	110132	-----TATTACAAlITACGAGAGTAAAAAATCTTAGCAGAGTGAAACATCTCATGTGANTAC	11

QY 59 aspglyglutlureuaspglmglylursPheValValaspgluasn---PheThrgluasn 77
 :: :: ||| ||||| :: ::|||
 Db 110132 ---TATTCAAAATTCGAAGATGAAAACATCTACGAAGATGAAAACATCTACTATGATTAAC 110076

b10075 TATGATCTTGACCATGCATTAAGCATTGGATAAAGCGAGCATGAT 1111:::

oy /8 tyrlenthraspcysgluglylysaspalagllysnalaalaglythrclgylaspluser 97

QY 98 AspGluValAspGluAsp 103
|||||

Db 110015 GACGATGTCGACGACGAC 109998

RESULT 10

LOCUS SFU15038 2255 bp mRNA linear TNY 06-JAN-1999

DEFINITION	Spodoptera frugiperda immunophilin FKBP46 mRNA, complete cds.
ACCESSION	U15038

```

VERSION
U15038.1  GI:595844

```

isomerase.

ORGANISM Spodoptera frugiperda

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

ORGANISM *Plasmodium falciparum*

Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GHGE
 Center clone name: CH230-68L8
 ----- Summary Statistics
 Sequencing vector: plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 135696 bases at least Q40
 Consensus quality: 141880 bases at least Q30
 Consensus quality: 146447 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 62 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 1 1057: contig of 1057 bp in length
 * 1058 1157: gap of unknown length
 * 1158 2631: contig of 1474 bp in length
 * 2632 2731: gap of unknown length
 * 2732 4125: contig of 1394 bp in length
 * 4126 4225: gap of unknown length
 * 4226 5385: contig of 1060 bp in length
 * 5386 6386: gap of unknown length
 * 6387 6486: contig of 1001 bp in length
 * 6487 7665: gap of unknown length
 * 7666 7765: contig of 1179 bp in length
 * 7766 8963: gap of unknown length
 * 8964 9063: contig of 1198 bp in length
 * 9064 10363: gap of unknown length
 * 10364 10463: contig of 1300 bp in length
 * 10464 11539: gap of unknown length
 * 11540 11639: contig of 1076 bp in length
 * 11640 12696: gap of unknown length
 * 12697 12796: contig of 1057 bp in length
 * 12797 14269: gap of unknown length
 * 14270 14369: contig of 1473 bp in length
 * 14370 16416: gap of unknown length
 * 16417 16516: contig of 2047 bp in length
 * 16517 17795: gap of unknown length
 * 17796 17895: contig of 1279 bp in length
 * 17896 19000: gap of unknown length
 * 19001 19100: contig of 1105 bp in length
 * 19101 21472: gap of unknown length
 * 21473 21572: contig of 2372 bp in length
 * 21573 22813: gap of unknown length
 * 22814 22913: contig of 1241 bp in length
 * 22914 24557: gap of unknown length
 * 24558 24657: contig of 1644 bp in length
 * 24658 25950: gap of unknown length
 * 25951 26050: contig of 1293 bp in length
 * 26051 27075: gap of unknown length
 * 27076 27175: contig of 1025 bp in length
 * 27176 28466: gap of unknown length
 * 28467 28567: contig of 1291 bp in length
 * 28567 30244: gap of unknown length
 * 30245 30344: contig of 1678 bp in length
 * 30345 32012: gap of unknown length
 * 32013 32112: contig of 1666 bp in length
 * 32113 34360: gap of unknown length
 * 34361 34460: contig of 2248 bp in length
 * 34461 36531: gap of unknown length
 * 36532 36631: contig of 2071 bp in length
 * 36632 38635: gap of unknown length
 * 38636 38735: contig of 2004 bp in length
 * 38736 39851: gap of unknown length
 * 39851 39851: contig of 1116 bp in length

39852 39951: gap of unknown length
 * 39952 41564: contig of 1613 bp in length
 * 41565 41664: gap of unknown length
 * 41666 43554: contig of 1890 bp in length
 * 43555 43654: gap of unknown length
 * 43656 45675: contig of 2021 bp in length
 * 45676 45775: gap of unknown length
 * 45776 47582: contig of 1807 bp in length
 * 47583 47682: gap of unknown length
 * 47683 49461: contig of 1779 bp in length
 * 49462 49561: gap of unknown length
 * 49562 51228: contig of 1667 bp in length
 * 51229 51328: gap of unknown length
 * 51329 53972: contig of 2644 bp in length
 * 53973 54072: gap of unknown length
 * 54073 55754: contig of 1682 bp in length
 * 55755 55854: gap of unknown length
 * 55855 58313: contig of 2459 bp in length
 * 58314 58414: gap of unknown length
 * 58414 60192: contig of 1779 bp in length
 * 60193 60292: gap of unknown length
 * 60293 63223: contig of 2931 bp in length
 * 63224 63323: gap of unknown length
 * 63324 65592: contig of 2269 bp in length
 * 65593 65692: gap of unknown length
 * 65693 67033: contig of 1341 bp in length
 * 67034 67133: gap of unknown length
 * 67134 69250: contig of 2117 bp in length
 * 69251 69350: gap of unknown length
 * 69351 71303: contig of 1953 bp in length
 * 71304 71403: gap of unknown length
 * 71404 74627: contig of 3224 bp in length
 * 74628 74727: gap of unknown length
 * 74728 77438: contig of 2711 bp in length
 * 77439 77538: gap of unknown length
 * 77539 79836: contig of 2298 bp in length
 * 79837 79936: gap of unknown length
 * 79937 81668: contig of 1732 bp in length
 * 81669 81768: gap of unknown length
 * 81769 84287: contig of 2518 bp in length
 * 84287 84386: gap of unknown length
 * 84387 88346: contig of 3960 bp in length
 * 88347 88446: gap of unknown length
 * 88447 92282: contig of 3836 bp in length
 * 92283 92382: gap of unknown length
 * 92383 97293: contig of 4910 bp in length
 * 97293 97393: gap of unknown length
 * 97393 100776: contig of 3384 bp in length
 * 100777 100876: gap of unknown length
 * 100877 104217: contig of 3341 bp in length
 * 104218 104317: gap of unknown length
 * 104318 107540: contig of 3223 bp in length
 * 107541 107640: gap of unknown length
 * 107641 113197: contig of 5557 bp in length
 * 113198 113297: gap of unknown length
 * 113298 118488: contig of 5191 bp in length

Alignment Scores:

Pred. No.: 821 Length: 178681
 Score: 80.50 Matches: 18
 Percent Similarity: 48.15% Conservatve: 8
 Best Local Similarity: 33.33% Mismatches: 9
 Query Match: 14.05% Indels: 19
 DB: 2 Gaps: 3

US-09-936-737a-2 (1-103) x AC098520 (1-178681)

QY 7 TtrpRpheryrAlaAsnArGLysTYrTrhAspPhesAspLysSerPhelLysSerSer 26

DB 37588 TGGAGTTCCTCAAAATCATATGTCCTC----- 37553

QY 27 AspleuaspGLucLysLysTrhCysPhelLysTrhGlutYrCysTrpLysValpneGlu 46

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

Hernandez, O., Hodgson, A., Hughes, M., Holloway, C., Hollins, B.,
 Homsl, F., Howard, S., Huber, J., Buliy, S., Hume, J., Jackson, L.E.,
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
 Kratochvic, U., Kuresh, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.
 Li, J., Li, Z., Lichtenarge, O., Lieu, C., Liu, J., Liu, W., Louisgeed, H.,
 Lozdo, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Mafshahari, M., Mapua, P., Martin, R., Mattindale, A., Martinez, E.,
 Massey, E., McElwain, E., McLeod, M.P., Meador, M., Mel, G., Metzker, M.,
 Mier, G., Miner, Z., Mitchell, T., Mohabdat, K., Morgan, M., Morris, S.,
 Moser, M., Neel, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokweto, S., Oguh, M., Okunuga, G.,
 Oragunye, L., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, J.,
 Rives, M., Rojas, A., Rojubo, K., I., Rolfe, M., Ruiz, S., Saveny, G.,
 Scherer, S., Scott, G., Shen, H., Shoshari, N., Sisson, L.,
 Sodergren, E., Sonalike, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Swatek, A., Tabort, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, T., Teifrod, B., Thomas, N., Thomas, S.,
 Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, O.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G., and Gibbs, R.

Direct Submission
 2 (bases 1 to 232370)
 Unpublished

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

Submitted Submission
 Submitted (29-NOV-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 232370)
 Worley, K.C.

Direct Submission
 Submitted (12-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 11, 2002 this sequence version replaced 413046745.

```

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GPDR
Center clone name: CH230-10M24
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990379
Consensus quality: 196541 bases at least Q40
Consensus quality: 199662 bases at least Q30
Consensus quality: 201786 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length.
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_drift_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 38 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 1404 1403: contig of 1403 bp in length
* 1504 1503: gap of unknown length
* 1504 2753: contig of 1250 bp in length
* 2754 2853: gap of unknown length
* 2854 4531: contig of 1678 bp in length
* 4532 4631: gap of unknown length
* 4632 5833: contig of 1192 bp in length
* 5824 5923: gap of unknown length

```

60783 a 51173 c 49242 g 60940 t 10232 others

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 25, 2003, 02:32:15 ; Search time 229 Seconds

Title: US-09-936-737a-2 (without alignments)
1012.908 Million cell updates/sec

Perfect score: 573
Sequence: 1 EEREDCWTFYANKRYTDFDK.....GKDAGNAGTGDESDVEDV 103

Scoring table:

BL0SUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame-p2n.model -DEV=x1h
-Q/cgnt_1/USPTO-spool/US0936737/runat_14022003.100732.2281/app.query.fasta.1.263
-DB=N_Geneseq_101002 -OFMT=fastfp -SUPFIX=ring -MINMATCH=0.1 -LOOPTCL=0
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptio -NORR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0936737 -CGCN_1.1.18 -etunat_14022003.100732.2281 -NCPU=6 -ICPU=3
-NO_XLPYX -NO_MAP -LARGEQUERY -NEG_SCORES=0 -NAIT -LONGLOG -DEV TIMECUT=120
-YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7
-YGAPOP=6 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N_Geneseq_101002: *
1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT: *
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT: *
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT: *
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT: *
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT: *
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT: *
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT: *
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT: *
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT: *
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT: *
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT: *
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT: *
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT: *
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT: *
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT: *
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT: *
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT: *
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT: *
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT: *
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT: *
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT: *
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001.DAT: *
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT: *
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT: *

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	573	100.0	375	AAA28180	Platelet binding 1
2	81.5	14.2	2264	AAO4441	S. frugiperda immu
3	80.5	14.0	4441	AAQ23682	leech antiplatelet
4	77	13.4	1107	ABL26101	Drosophila melanog
5	76.5	13.4	587	ABQ47468	Oligonucleotide fo
6	76.5	13.4	587	ABQ47468	Oligonucleotide fo
7	76	13.3	2496	ABQ47469	Borrelia burgdorfe
8	75.5	13.2	1399	AAK20297	Human interleukin-
9	75.5	13.2	1399	AAK20297	Human interleukin-
10	75.5	13.2	1402	AAK20297	Human interleukin-
11	75.5	13.2	1402	AAK20297	Human interleukin-
12	75.5	13.2	1402	AAK20297	Human interleukin-
13	75.5	13.2	1402	AAK20297	Human interleukin-
14	74.5	13.1	2562	ABV25467	Drosophila melanog
15	74.5	13.0	2562	ABV25467	Drosophila melanog
16	73.5	12.8	1116	AAK84453	Human prostate exp
17	73.5	12.8	1116	AAK84453	Human prostate exp
18	73.5	12.8	1116	AAK84453	Human prostate exp
19	72	12.6	27884	AAK77781	Human immune/haema
20	71.5	12.5	762	ABN70859	Drosophila melanog
21	71.5	12.5	762	ABN70859	Drosophila melanog
22	71.5	12.5	762	ABN70859	Drosophila melanog
23	71.5	12.5	762	ABN70859	Drosophila melanog
24	71.5	12.5	762	ABN70859	Drosophila melanog
25	71.5	12.5	762	ABN70859	Drosophila melanog
26	71.5	12.5	762	ABN70859	Drosophila melanog
27	71.5	12.5	762	ABN70859	Drosophila melanog
28	71.5	12.5	762	ABN70859	Drosophila melanog
29	71.5	12.5	762	ABN70859	Drosophila melanog
30	71.5	12.5	762	ABN70859	Drosophila melanog
31	71	12.4	215561	ABN71527	Human CDNA sequenc
32	71	12.4	2229	ABL22190	Drosophila melanog
33	70.5	12.3	9842	AAK20259	Borrelia burgdorfe
34	70.5	12.3	661	AAH1808	Human differential
35	70.5	12.3	661	AAH1808	Human differential
36	70.5	12.3	661	AAH1808	Human differential
37	70.5	12.3	1370	ABV24960	Drosophila melanog
38	70	12.2	540	ABV25121	Human prostate exp
39	70	12.2	2010	AAV88360	Human prostate exp
40	70	12.2	2010	AAV88360	Human prostate exp
41	70	12.2	2010	AAV88360	Human prostate exp
42	70	12.2	2010	AAV88360	Human prostate exp
43	70	12.2	2010	AAV88360	Human prostate exp
44	69.5	12.1	344	AAK8317	Human digestive sy
45	69.5	12.1	344	AAK8317	Human digestive sy

ALIGNMENTS

RESULT 1
ID AAA28180 standard; CDNA: 375 BP.
XX
AC AAA28180:
XX
29-JAN-2001 (first entry)
XX
DE Platelet binding inhibitor protein Saratin encoding CDNA sequence.
XX
KW Saratin: medicinal leech, thromboembolic disease; intraocular lens;
KW collagen-dependent platelet adhesion inhibition;
KW posterior capsule opacification; ss.
XX
OS Hirudo medicinalis.
XX
FH
FT

SAWE

Key Location/Qualifiers
CDS 64..375

```

FT      /*tag= a "Saratin"
FT      /product= "Saratin"
FT      /note= "Platelet adhesion inhibitor protein"
XX      WO200056885-A1.
XX      28-SEP-2000.
XX      10-MAR-2000; 2000WO-EP02117.
XX      18-MAR-1999; 99EP-010530.
XX      12-MAY-1999; 99EP-0109503.
XX      (MERE ) MERCK PATENT GMBH.
XX      Stittmatter W, Guessow D, Hofmann U, Hemberger J, Fotev Z;
XX      Scheuble B;
XX      WPI: 2000-611629/58.
XX      P-PSDB; AAY94746.
XX      New Saratin polypeptide and gene isolated from Hirudo medicinalis for
XX      blocking platelet adhesion, especially useful for treating or
XX      preventing thrombotic diseases, or for manufacturing a medicament for
XX      thromboembolic diseases
XX
XX      Claim 4; Page 40; 46pp; English.
XX
XX      Saratin is a protein isolated from the saliva of the medicinal leech
XX      Hirudo medicinalis. Saratin is an inhibitor of collagen-dependent
XX      platelet adhesion. The invention includes polynucleotide sequences, a
XX      encoding Saratin, an expression vector comprising the DNA sequence, a
XX      host cell transformed with the expression vector, antibodies
XX      immunospecific for Saratin, and methods for identifying Saratin agonists,
XX      or antagonists. Saratin is useful for treating thromboembolic diseases,
XX      and for manufacturing a medicament for treating thromboembolic diseases.
XX      It is useful for preventing thrombotic diseases. Saratin is also useful
XX      for coating artificial surfaces, since use of Saratin renders them
XX      non-adhesive for cells and prevents the activation of cells. It may also
XX      be used for coating natural collagen surfaces. Furthermore, Saratin is
XX      useful for modifying intraocular lenses in order to lessen the
XX      thrombogenicity of the lens material, for contacting the lens surface, or
XX      for covalent crosslinking to modify the lens material. The lens material
XX      is used for refractive anterior or posterior chamber ocular implants,
XX      which may be implanted into the eye. This new type of coating avoids
XX      problems contributed by stimulated cell growth. In combination with other
XX      medicaments that are for instance conferring cell death, Saratin coating
XX      helps to completely overcome posterior capsule opacification. The
XX      antibody immunospecific for Saratin, as well as Saratin itself, are
XX      useful for measuring samples derived from host cell cultures or from a
XX      treated subject. The present sequence represents cDNA encoding the
XX      Saratin protein.
XX
XX      Sequence 375 BP; 123 A; 62 C; 85 G; 105 T; 0 other:
XX
XX      Alignment Scores:
XX      Pred. No.: 7 28e-64 Length: 375
XX      Score: 573.00 Matches: 103
XX      Percent Similarity: 100.008 Conservative: 0
XX      Best Local Similarity: 100.008 Mismatches: 0
XX      Query Match: 100.008 Indels: 0
XX      DB: 21 Gaps: 0
XX
XX      US-09-936-737a-2 (1-103) x AAA28180 (1-375)
XX
QY      1 GIUGIUAAGGUAAGCAGTGTGTTGACGCGAACAAGAAATGTACAGCTTCGATGATAA 20
DB      64 GAAAGACGTAAGATGTTGACGCGAACAAGAAATGTACAGCTTCGATGATAA 123
QY      21 SerpHeuyslvysSerSerAspLeuAspGluCysLysLysThrCysPheLysThrGluTyr 40
DB      124 TCTTTTAAGAAAGTCCCTGATCTTGACGAATGCAAAAACATGTTTCAGACGAGATAC 163

```

```

QY      41 CysTyrTlleValPheGluAspThrValAsnLysGluCysTyrTyrAsnValValAspGly 60
DB      184 TGCATCATCGTTTGGAGACGTCACAGCAAGCAATGTACTACAAATCGCTTATGCT 243
QY      61 GluGluLeuAspGlnGlnLysPheValValAspGluAsnPheThrGluAsnTyrLeuThr 80
DB      244 GAAAGATTAGACCAAGAAAAATTTGTTGCGACGAAACTCAGCGAAAAATTTATTTGACA 303
QY      81 AspCysGluGlyLysAspAlaGlyAsnAlaAlaGlyThrGlyAspGluSerAspGluVal 100
DB      304 GACTCGCGAGGCTAAAGATGACGATTAATGCGCAGGTACAGGTGACGATGATGAGTT 363
QY      101 AspGluAsp 103
DB      364 GATGAAAGAT 372
XX
XX      RESULT 2
XX      AAX04441
XX      AAX04441 standard; cDNA; 2284 BP.
XX
XX      AAX04441;
XX      27-APR-1999 (first entry)
XX      S. frugiperda immunophilin FKBP46 cDNA.
XX      Immunophilin; moth; insect cell; nuclear; immunosuppression; drug;
XX      transplant; tissue graft; ds.
XX      Spodoptera frugiperda.
XX
XX      Key Location/Qualifiers
XX      CDS 160..1398
XX      FT /*tag= a
XX      FT /product= "immunophilin FKBP46"
XX      FT polyA_signal 1764..1769
XX      FT polyA_signal 1850..1855
XX      FT polyA_signal /*tag= b
XX      FT polyA_signal 1868..1873
XX      FT polyA_signal /*tag= c
XX      FT polyA_signal 2052..2057
XX      FT polyA_signal /*tag= d
XX      FT polyA_signal 2237..2242
XX      FT polyA_signal /*tag= e
XX      FT polyA_signal /*tag= f
XX
XX      US861498-A.
XX      19-JAN-1999.
XX      31-OCT-1996; 96US-0741134.
XX      01-NOV-1995; 95US-0007163.
XX      31-OCT-1996; 96US-0741134.
XX      (UJUE-) UNIV JEFFERSON THOMAS.
XX
XX      Alnemri ES, Fernandes-Alnemri T, Litwack G;
XX      WPI: 1999-130433/11.
XX      P-PSDB; AAW68010.
XX
XX      Isolated immunophilin FKBP46 nucleic acids - useful for developing
XX      products for the study and identification of immunosuppressive
XX      agents for treating e.g. transplant and tissue graft patients
XX
XX      Claim 2; Fig 1; 21pp; English.
XX
XX      This sequence represents the cDNA encoding the immunophilin FKBP46 from
XX      the moth Spodoptera frugiperda. FKBP46 is a novel insect cell nuclear
XX      immunophilin, which can be used to study and identify additional
XX      immunosuppressive drugs that bind to it. Such immunosuppressive agents
XX      can be used in treating transplant and tissue graft patients.

```


01-SEP-2001; 2001WO-EP10074

12-JUL-2002 (first entry)

```

Db      226 GTTAAAGAGAAATATGTTGCTTTCGTTCGGCGTCGACGATTGGGTAACGCTTCGCTAGT 16
QY      94 GLYAspGlnSer-----AspGluValAspGluAsp 103
      |||  ::|  |||  ||| ||| ||| |||
Db      166 AGCGATTGACGCGGGGTTTGACGACGATTAATTACGATGAGAGAT 125
RESULT 7
AAAX20297/c
ID      AAX20297 standard; DNA; 2496 BP.
XX
XX      AAX20297;
XX
XX      04-MAY-1999 (first entry);
DE      Borrelia burgdorferi polynucleotide sequence #50.
XX
XX
KW      Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
KW      epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
XX      infection; diagnosis; characterisation; detection; ds.
XX
OS      Borrelia burgdorferi.
XX
XX      MO9858943-X1.
XX
XX      30-DEC-1998.
XX
XX      18-JUN-1998; 98MO-US12764.
XX
XX      03-SEP-1997; 97US-0057483.
XX      PR 20-JUN-1997; 97US-0050359.
XX      PR 22-JUL-1997; 97US-0053344.
XX      PR 22-JUL-1997; 97US-0053377.
XX
XX      (HUMA-) HUMAN GENOME SCI INC.
XX      (MEDIM-) MEDIMUNE INC.
XX
XX      Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO;
XX      White OR;
XX
XX      WPI; 1999-081217/07.
XX
XX
XX      New isolated Borrelia burgdorferi nucleic acids - used to develop
XX      products for the detection, diagnosis, characterisation, prevention
XX      and therapy of infections, particularly Lyme disease
XX
XX      Claim 1; Page 1026-1028; 1128pp; English.
XX
XX      AAX20248 to AAX20402 represent polynucleotide sequences isolated from
XX      CC Borrelia burgdorferi (Bb). Products derived from Bb can be used for
XX      CC the detection, diagnosis, characterisation, prevention and therapy of
XX      CC Bb infections, e.g. Lyme disease. They can also be used for the
XX      CC production of biosynthetic products, e.g. enzymes. Borrelia belongs
XX      CC to a family of motile, spiral-shaped bacteria called Spirochetes.
XX      CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and
XX      CC endemic relapsing fever, and Lyme borreliosis, more commonly known as
XX      CC Lyme disease.
XX
XX
SQ      Sequence 2496 BP; 871 A; 348 C; 235 G; 1029 T; 13 other;
Alignment Scores:
Pred. No.: 13.3 length: 2496
Score: 76.00 Matches: 18
Percent Similarity: 48.61% Conservative: 17
Best Local Similarity: 25.00% Mismatches: 25
Query Match: 13.26% Indels: 12
DB: 20 Gaps: 3
US-09-936-737A-2 (1-103) x AAX20297 (1-2496)
QY      10 TYRAlaAsnArgLysTYRThrAspPheAspLysSerPheLysLysSerSerAspLeuAsp 29
      |||  ::|  |||  ||| ||| ||| |||
Db      344 TATGACGAGAAAAAAGATATATTGATTTGAT-----TTTCTTAATAATTAAGAAAAAGAA 291

```

QY	30	GLHCysLySLySLyTThnCpSpheLySLyThGluTyrCysSfYrILLeValPheGluAspPhrVal	49
Db	290	GAA-----TATPAAGAGCAAGAAATGGTTATTTATTTATTTAGCTATCTTATA	24
QY	50	AsnLySLyGluCysTfYrYAsnVal-----ValAspLySLyGluLeuAsp	64
Db	245	AGAAAGAAAGATACCTAAARCAAGATATTTATATKACAGTTGAAAGARCAAAATATCY	18
QY	65	GLnGluLySLyPheValValAspSLyAsnPherThcGlu	76
Db	185	AAAGAAAGAAAGAKATATATTCAAAACCTCTTAGAGAG	150
RESULT 8			
ID	AAQ79969	standard; CDNA: 1399 BP.	
XX	AAQ79969;		
DT	12-SEP-1995	(first entry)	
DE	Human Interleukin-1 beta converting enzyme homolog (Ice-4).		
KM	Human Interleukin-1 beta converting enzyme homolog; Ice-4;		
OS	oncogene bcl-2; programmed cell death; cancer treatment; ss.		
XX	Homo sapiens.		
XX	Key	Location/Qualifiers	
FT	mat.peptide	3..1297	
FT		/*tag= a	
PN	MO9500160-A.		
XX	05-JAN-1995.		
XX	10-JUN-1994;	94WO-US06630.	
XX	24-JUN-1993;	93US-0080850.	
XX	(GEHO) GEN HOSPITAL CORP.		
PI	Mitura M, Yuan J;		
DR	WPI: 1995-051742/07.		
DR	P-PSDB; AAR66769.		
XX	Promoting or preventing programmed cell death in vertebrate cells		
PT	- by inhibiting the activity of Interleukin-1 beta converting		
PT	enzyme.		
PS	Claim 21; Fig 16; 116pp; English.		
XX	AAQ79969 encodes AAR66769 human interleukin-1 beta converting enzyme		
CC	homolog (Ice-4), increasing Ice-4s enzymatic activity can		
CC	promote the programmed cell death of cancer cells (pref. those		
CC	overexpressing the bcl-2 oncogene), this can be used as the basis		
CC	of a new cancer treatment. Alternatively by reducing Ice-4s enzymati		
CC	activity programmed cell death can be inhibited, this may be useful		
CC	in the development of new cell lines which remain viable in culture		
CC	for extended or indefinite periods, independant of growth factors.		
XX	Sequence 1399 BP; 440 A; 270 C; 316 G; 373 T; 0 other;		
XX	SD		
Alignment Scores:			
Pred. No.:	7.29	Length:	1399
Score:	75.50	Matches:	26
Percent Similarity:	54.79%	Conservative:	14
Best Local Similarity:	35.62%	Mismatches:	14
Query Match:	13.18%	Indels:	19
DB:	16	Gaps:	6

Oy	44	VaIphgGluAspPhrVaIAsnIsgLusCysTrYtYrAsnAlValAspLygluIleu	63
Db	120	GTTTTTGAATGACCTGGTGAGAAC-----AATGTTTTAATGAAATGACTG	167
Oy	64	-----AspIgcIuIuIsPheVaIAsp-----GluAsnPheThrluAsn	77
Db	168	CTCAAAAATAGGGGAAGTCGACAGTTTCATCTCGAACAAGCTGAGATCTGGTTAGAAC	227
Oy	78	TyrIeu-----ThrAspCysIsgLyllys---AspIadIysmIadIaclyThrcly	94
Db	228	TTCTTTAGAGAAAAACAGCATGGCAGAAAATAATTCTGTGGCCACATTCACATTCACAG	287
Oy	95	AspIcu-----SerAspIcuVaIasp	101
Db	288	GAACACGCTGAGTTTACAAATTTCTTAATGATAGAGATGAT	326
RESULT 9			
ID	AAT31554		
XX	AAT31554 standard; cDNA; 1399 BP.		
XX	AAT31554;		
DI	25-SEP-1996 (first entry)		
DE	Ich-3 cDNA.		
XX	Ich-3; murine ICE-ced-3 homologue; programmed cell death;		
KW	apoptosis; interleukin-1 beta converting enzyme; gene therapy; ss.		
XX	Mus sp.		
OS			
FH	Key	Location/Qualifiers	
FT	CDS	3..1301	
FT		/tag= a	
PN	MO9620721-A1.		
PD	11-JUL-1996.		
XX	PF	04-JAN-1996; 96WO-US00177.	
XX	PR	04-JAN-1995; 95US-0368704.	
XX	(GEHO) GEN HOSPITAL CORP.		
PA	Miura M, Yuan J;		
PI	DR	WPJ; 1996-333763/33.	
XX	P-PSDB; AAR98464.		
PT	Preventing or promoting programmed cell death in vertebrate cells		
PT	comprises inhibiting or increasing the activity of		
PT	interleukin-1-beta converting enzyme, or altering expression of		
PT	other related genes		
XX	Claim 21; Fig 14; 127pp; English.		
XX	A novel murine cell death gene (AAT31554), designated Ich-3, was		
CC	identified as a member of the ICE/ced-3 family. It codes for a		
CC	protein (AAR98464) that shows homology to mouse interleukin-1 beta		
CC	converting enzyme (ICE), mch-2 (see also AAR98461) and human Ich-1		
CC	(see also AAR98462-63). Full-length Ich-3 cDNA was isolated from a		
CC	mouse thymus cDNA library. Expression of Ich-3 can be used in		
CC	methods of regulating cell death, while antisense sequences may be		
CC	used to prevent programmed cell death.		
XX	Sequence 1399 BP; 440 A; 270 C; 316 G; 373 T; 0 other;		
SQ			
Alignment Scores:			
Pred. No.:	7.29	Length:	1399
Score:	75.50	Matches:	26
Percent Similarity:	54.79%	Conservative:	14
Best Local Similarity:	35.62%	Mismatches:	14

```

Query Match:      13.18%      Indels:      19
DB:               17          Gaps:          6
US-09-936-737a-2 (1-103) x AA332645 (1-1399)

Qy  44 ValPhegluAspThrValAsnLysGluCysTyrTyrAsnValValAspGluGluLeu 63
    |||||:|||||  |||  |||  |||||:|||||:|||||:|||||:|||||
Db  120 GTTTTGTATGACCTGGTGGAGAAAG-----AACTTTTAATGAGATGAGTTG 167
Qy  64 -----AspGluGluLysPheValValAsp-----GluAsnDheThrgLusn 77
    :|||:|||||:|||||:|||||:|||||:|||||:|||||
Db  168 CTCAAATAGGGGAAAGCTGCGATTCATCTGAAACAAGCTGCAATCTGTTGAGAAC 227
Qy  78 TyrLeu-----ThraspCysGluGlyLys---AspAlaGlyAsnAlaIaGlyThrgly 94
    :|||:|||||  |||||  |||||  |||||:|||||:|||||:|||||
Db  228 TTCTTAGAGAAACAGACATGCGAGAAATTTGCTGCGCCACATTCCTCCATTCACG 287
Qy  95 AspGlu-----SerAspGluValAsp 101
    :|||:|||||  |||  |||
Db  288 GAACAGCTGAGTTTACAAATTTTCTTAATGATGAGATGAT 326

RESULT 10
AA332645
ID  AA332645 standard; cDNA; 1402 BP.
XX
AC  AA332645;
XX
DT  21-JAN-2000 (first entry)
XX
DE  Murine caspase-12 splice variant, caspase-12L cDNA.
XX
KW  Caspase; splice variant; truncated; programmed cell death; apoptosis;
KW  regulation; proteolytic cascade; malignant condition; cancer;
KW  premalignant condition; solid tumour; lymphoma;
KW  chronic lymphocytic leukaemia; prostatic hypertrophy;
KW  preneoplastic liver focus; chemotherapy resistance; autoimmune disease;
KW
XX
OS  Mus sp.
XX
FH  Key
FT  CDS
FT  1..1050      Location/Qualifiers
FT  /*tag= a      /product= "Murine caspase-12 splice variant,
FT  /product=     caspase-12L"
FT  exon
FT  1..93
FT  /*tag= b      /number= 1
FT  94..303
FT  /*tag= c      /number= 2
FT  304..1402
FT  /*tag= d      complement (918..943)
FT  primer_bind   /bound_motety= "Degenerate upstream PCR primer AA332648"
FT  1281..1300
FT  /*tag= f
FT  /bound_motety= "Degenerate downstream PCR primer
FT  AA332649"
XX
FN  W09952925-A1.
XX
PD  21-OCT-1999.
XX
PE  14-APR-1999; 99MO-US08064.
XX
PR  16-APR-1998; 98US-0081962.
XX  (GEHO ) GEN HOSPITAL CORP.
XX
PI  Yuan J, Morishima N;
XX

```

```

DR  WPI: 1999-620369/53.
DR  P-PSDB: AA50109.
XX
PT  New nucleic acid encoding the short form of caspase-12, used e.g. for
PT  treating tumors
PS  Example 1; Fig 1; 68pp; English.
XX
CC  This sequence represents a cDNA encoding murine caspase-12L, a splice
CC  variant of caspase-12. Caspases are a family of proteins involved in
CC  the regulation of apoptosis and are synthesised as proforms which are
CC  activated via cleavage after specific Asp residues. Mammalian cells
CC  express several caspases, and it is thought that these act in a
CC  proteolytic cascade to cause programmed cell death. Nucleic acids
CC  encoding caspase-12S (AA332644) or truncated forms of caspase-12L
CC  (AA332646, AA332647) are used for production, recombinantly or in vivo,
CC  of caspase-12 polypeptides which induce programmed cell death. This is
CC  particularly useful for treating (pre)malignant conditions (e.g., solid
CC  tumours, B cell lymphoma, chronic lymphocytic leukaemia, prostatic
CC  hypertrophy, preneoplastic liver foci and resistance to chemotherapy), or
CC  autoimmune diseases. The caspase-12 proteins can also be used to raise
CC  specific antibodies (for example, to determine gene expression and to
CC  screen expression libraries) or as molecular weight markers. Fragments
CC  of caspase-12-encoding nucleic acids can be used as probes to isolate
CC  the caspase-12 gene (and its allelic variants); in fluorescent in situ
CC  hybridisation for chromosomal location of the caspase-12 gene, and for
CC  Northern blotting to determine caspase-12 mRNA expression in tissues.
XX
SQ  Sequence 1402 BP; 442 A; 270 C; 315 G; 375 T; 0 other;

Alignment Scores:
Pred. No.:      7.31      Length:      1402
Score:          75.50      Matches:      26
Percent Similarity: 54.79%      Conservative: 14
Best Local Similarity: 35.62%      Mismatches:  14
Query Match:    13.18%      Indels:      19
DB:             20          Gaps:       6

US-09-936-737a-2 (1-103) x AA332645 (1-1402)

Qy  44 ValPhegluAspThrValAsnLysGluCysTyrTyrAsnValValAspGluGluLeu 63
    |||||:|||||  |||  |||  |||||:|||||:|||||:|||||
Db  120 GTTTTGTATGACCTGGTGGAGAAAG-----AACTTTTAATGAGATGAGTTA 167
Qy  64 -----AspGluGluLysPheValValAsp-----GluAsnDheThrgLusn 77
    :|||:|||||:|||||:|||||:|||||:|||||:|||||
Db  168 CTCAAATAGGGGAAAGCTGCGATTCATCTGAAACAAGCTGCAATCTGTTGAGAAC 227
Qy  78 TyrLeu-----ThraspCysGluGlyLys---AspAlaGlyAsnAlaIaGlyThrgly 94
    :|||:|||||  |||||  |||||  |||||:|||||:|||||:|||||
Db  228 TTCTTAGAGAAACAGACATGCGAGAAATTTGCTGCGCCACATTCCTCCATTCACG 287
Qy  95 AspGlu-----SerAspGluValAsp 101
    :|||:|||||  |||  |||
Db  288 GAACAGCTGAGTTTACAAATTTTCTTAATGATGAGATGAT 326

RESULT 11
AA332643
ID  AA332643 standard; cDNA; 1402 BP.
XX
AC  AA332643;
XX
DT  09-FEB-2001 (first entry)
XX
DE  Mouse Ice-4 coding sequence.
XX
KW  ced-3; virally induced cell death; apoptosis; gene therapy; neural;
KW  muscular degenerative disease; myocardial infarction; stroke; aging;
KW  interleukin-1beta converting enzyme; ICE; mouse;
XX  Ice-ced 3 homologue; Ich, ss.
XX
OS  Mus sp.
XX

```

Key Location/Qualifiers
 CDS 45..1304
 /*tag= a
 /product= "Mouse Ice-4"

US6083735-A.
 04-JUL-2000.
 10-JUN-1994; 94US-0258287.
 24-JUN-1993; 93US-0080850.
 (GENO) GEN HOSPITAL CORP.
 Yuan J, Miura M;
 WPI; 2000-464343/40.
 P-PSDB; AAB14259.

New human Ich-1L and Ich-1S proteins for negative and positive regulation of programmed cell death and for developing therapeutic methods for diseases and conditions characterized by cell death, e.g. myocardial infarction or stroke

Disclosure; Fig 16; 121pp; English.

The present sequence is the mouse Ice-4 protein coding sequence. The protein sequence of the present sequence was used in a sequence homology comparison with the protein sequences of human Ice-3 homology (Ich-1) (AAB14253), murine Interleukin-1beta converting enzyme (mICE) (AAB14249), murine ICE2 (AAB14252) and C. elegans ced-3 (AAB14246). The present sequence is a member of a family of genes involved in programmed cell death (apoptosis). Ich-1 may play an important role in both the positive and negative regulation of apoptosis. The Ich gene may be used in gene therapy in disorders characterised by cell death e.g. neural and muscular degenerative diseases, myocardial infarction, stroke, virally induced cell death and aging.

Sequence 1402 BP; 442 A; 270 C; 315 G; 375 T; 0 other;

Alignment Scores:
 Pred. No.: 7.31 Length: 1402
 Score: 75.50 Matches: 26
 Percent Similarity: 54.79% Conservative: 14
 Best Local Similarity: 35.62% Mismatches: 14
 Query Match: 13.18% Indels: 19
 DB: 21 Gaps: 6

US-09-936-737A-2 (1-103) x AAA72843 (1-1402)

OY 44 ValPheGluAspThrValAsnLysGluCysTyrTyrAsnValValAspGluGluLeu 63
 |||||
 DB 120 GTTTTGTGACGCTGTGAGAG-----AATCTTTAAATGAGATGAGTGA 167
 OY 64 -----AspGluGluLysPheValValAsp-----GluAsnPhetrGluAsn 77
 |||||
 DB 168 CTCGAAATATAGGGAAGGCGAGCTTCATCTGTAACAGCTGAGATCTGTTGAGAAC 227
 OY 78 TyrLeu-----ThrAspCysGluGluLys-----AspAlaGluAsnAlaIleGlyThrGly 94
 |||||
 DB 228 TTCTTAGAGAAACAGACATGCGAGAAATATTTCTGCGCCCATTTGCCAATTCGCCAG 287
 OY 95 AspGlu-----SerAspGluValAsp 101
 |||||
 DB 288 GACAGCTGAGTTTACATTTCTTAATCATGAGATGAT 326

RESULT 12
 ABL10668
 ID ABL10668 standard; cDNA; 4453 BP.
 XX ABL10668;
 AC
 XX

DT 26-MAR-2002 (first entry)
 XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 26486.
 DE Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.
 KW Drosophila melanogaster.
 OS
 PN WO200171042-A2.
 PD 27-SEP-2001.
 XX 23-MAR-2001; 2001WO-US09231.
 PF 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX (PEKE) PE CORP NY.
 PA Venter JC, Adams M, Li FWD, Myers EW;
 PI WPI; 2001-656860/75.
 XX P-PSDB; ABB6565.
 DR
 XX New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -
 PT Claim 1; SEQ ID NO 26486; 21pp + Sequence Listing; English.
 PS
 XX The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (AB16176-AB16175), expressed DNA sequences (AB16176-AB16175) and the encoded proteins (AB16176-AB16175).
 CC The sequence data for this patent did not form part of the printed CC specification, but was obtained in electronic format directly from WIPO CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 4453 BP; 1232 A; 1020 C; 1060 G; 1141 T; 0 other;

Alignment Scores:
 Pred. No.: 32.4 Length: 4453
 Score: 75.50 Matches: 30
 Percent Similarity: 48.91% Conservative: 15
 Best Local Similarity: 32.61% Mismatches: 36
 Query Match: 13.18% Indels: 11
 DB: 23 Gaps: 6

US-09-936-737A-2 (1-103) x ABL10668 (1-4453)

OY 4 GluAspCysTrpThrPheTyrAlaAsnArgLysTyrThrAspPheAspLysSerPheLys 23
 |||||
 DB 2124 GAGCAGGCCCGACGACGCTCTATCGCAGCAGCAACCAACA-----AAAAGTTTGAG 2174
 OY 24 LysSerSerAspLeuAspGluCysLysLysThrCysPheLysThrGluTrpCys----- 41
 |||||
 DB 2175 GCTGCTGAGATCTTCGCGTTTACTTCCAAATATATGCTT-----CAAGTATGTAATA 2228
 OY 42 ---TyrIleValIlePheGluAspThrValAsnLysGluCysTyrTyrAsnValValAspGly 60
 |||||
 DB 2229 GTTATATATAGTA---GAATTTTCATTAATGTAAGTAATGCTTGAAGAAATGCTGTAA 2285
 OY 61 GluGluLeuAspGluGluLysPheValValAspGluAsnPhetrGluAsnTyrIleuThr 80
 |||||
 DB 2286 TTTGAAGCCAGCTGAGACCGCTGCTCAGAGAGATGTTAAAGATAT---CTCACC 2342
 OY 81 AspCysGluGluLys-----AspAlaGluAsnAlaIle 91
 |||||

Db 2343 GATAGTACTACAAATATCCGATTTAATGCCGCG 2378
 RESULT 13
 AAK94875
 ID AAK94875 standard; cDNA; 2149 BP.
 XX
 AC AAK94875;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human full-length cDNA, SEQ ID NO: 4066.
 XX
 KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
 XX
 OS Homo sapiens.
 XX
 PN EPI30094-A2.
 XX
 PD 05-SEP-2001.
 XX
 PF 07-JUL-2000; 2000EP-0114089.
 XX
 PR 08-JUL-1999; 99JP-0194486.
 PR 11-JAN-2000; 2000JP-0118774.
 PR 02-MAY-2000; 2000JP-0183765.
 XX
 PA (HELI-) HELIX RES INST.
 PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y,
 PI Wakematsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 XX WPI: 2001-524255/58.
 DR P-PSDB; AAM93913.
 XX
 PT 830 Primers useful for synthesizing full length cDNA clones and their
 PT use in genetic manipulation
 XX
 PS Claim 8; SEQ ID NO 4066; 1380bp + sequence listing; English.
 XX
 CC The invention relates to primers for synthesizing full length cDNA
 CC clones; 830 cDNA molecules encoding a human protein have been
 CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
 CC molecules have been determined. Primers for synthesizing the full length
 CC cDNA are useful for clarifying the function of the protein encoded by
 CC the cDNA. The full length clones were obtained by construction of full
 CC length enriched cDNA libraries that were synthesised by the oligo-capping
 CC method. The primers enable the production of the full length cDNA easily
 CC without any special methods. The present sequence is a full length
 CC human cDNA of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in CD-ROM format directly from EPO.
 XX
 SO Sequence 2149 BP; 659 A; 415 C; 531 G; 544 T; 0 other;

Alignment Scores:
 Pred. No.: 14.7 Length: 2149
 Score: 75.00 Matches: 27
 Percent Similarity: 41.51% Conservative: 17
 Best Local Similarity: 25.47% Mismatches: 35
 Query Match: 13.09% Indels: 27
 DB: 22 Gaps: 4

US-09-936-737A-2 (1-103) x AAK94875 (1-2149)

QY 2 GUARGGLUASPCYSTIRPHRPHETRY-----AlaAsnArgIysTyrThrAspPhe 18
 Db 650 AAAAGGAAGACGCTGGTATTTTCCCTTGTCTGTAAGTTAAATATGCGATTTC 709
 QY 19 AsplysSerPhe-----LysIysSerSer 26
 Db 710 TCCTATCATTTTGGATTAATTCCTGGAAGCATTTCTAGAGAAATTAACCTTCACCTCT 769
 QY 27 AspleuAspGluCysLysIysThrCysPheLysThrGluTyrCysTyrIle-ValPheG1 46

Db 770 TCCTGCACAGAGCTTAATTGAAAAATTTTCCGAAAGAGTAT--TACATTGATGATGTC 826
 QY 46 uAspThrValAsnLysGluCysTyrTrpAsnValValAspGlyGluGluLeuAspGlnG1 66
 Db 827 AGAGTCACGCAATCACTGCTGATTTATGTTTCATTCACAAAGGTGAAGATGAGCAACA 886
 QY 66 uLys-----PheValValAspGluAsnPhetHrG1 76
 Db 887 AAAAAAAGAACCGATGATTTTGTGACAAAGGCTATTCTGTTGAGTCACTTCACGTGCG 946
 QY 76 uAsnTyrLeuThrAsp 81
 Db 947 TGAATGTTAGAAGAT 962

RESULT 14
 ABV25467
 ID ABV25467 standard; cDNA; 2562 BP.
 XX
 AC ABV25467;
 XX
 DT 16-SEP-2002 (first entry)
 XX
 DE Human prostate expression marker cDNA 25458.
 XX
 KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 KW pharmacogenomic marker; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200160860-A2.
 XX
 PD 23-AUG-2001.
 XX
 PF 20-FEB-2001; 2001WO-US05171.
 XX
 PR 17-FEB-2000; 2000US-183319P.
 PR 16-MAR-2000; 2000US-189862P.
 PR 25-MAY-2000; 2000US-207454P.
 PR 09-JUN-2000; 2000US-211314P.
 PR 18-JUL-2000; 2000US-219007P.
 PR 13-DEC-2000; 2000US-255281P.
 XX
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 PI Schlegel R, Endege WO, Monahan JR;
 XX WPI: 2001-662795/76.
 DR
 PT Novel isolated nucleic acid molecule associated with cancerous state of
 PT prostate cells and correlating with presence of prostate cancer, useful
 PT for detecting presence of prostate cancer, stage of prostate cancer -
 XX
 PS Claim 1; Page 5054; 11750bp; English.
 XX
 CC The invention relates to an isolated nucleic acid molecule (I) comprising
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
 CC specification or its complement. (I) is useful for:
 CC (a) assessing whether a patient is afflicted with prostate cancer;
 CC (b) monitoring the progression of prostate cancer in a patient;
 CC (c) assessing the efficacy of a test compound to inhibit prostate
 CC cancer in a patient;
 CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
 CC in a patient;
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;
 CC (f) assessing the prostate cell carcinogenic potential of a compound;
 CC (g) determining whether prostate cancer has metastasized in a patient;
 CC (h) assessing the aggressiveness or indolence of prostate cancer in a
 CC patient;
 CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.
 XX
 SO Sequence 2562 BP; 970 A; 365 C; 641 G; 576 T; 10 other;

Alignment Scores:
Pred. No.: 21.3 Length: 2562
Score: 74.50 Matches: 34
Percent Similarity: 45.76% Conservative: 20
Best Local Similarity: 28.81% Mismatches: 44
Query Match: 13.00% Indels: 20
DB: 23 Gaps: 7

US-09-936-737a-2 (1-103) x ABV25467 (1-2562)

QY 2 GLUARGGLAASPCYSTRPHRYR-----AlaasnrglystyrthrAsp-Ph 18
Db 657 GAAAAAATGTTCTCAACATCTCAGACCTCTCTCGAAGAAAAACAAGACACTT 716
QY 18 eaaplySerPheLysLys-----SerSerAspleuaspGlyCysLysLysThCysPh 36
Db 717 AGACTCAGCAGCCTCGAATCTGCAATCTCCAGAAATCGAGTGTTCAGACA----- 771
QY 36 elysThglutyrCysTyrIleVal---PhegluAspThrValasnLysglucysTyrTy 55
Db 772 -AGAGAGAAATCAATCACTGCTTCACTCATATGACAGACAGCAGTGTATGA 830
QY 55 rasnValaspGlygluGluLeuaspGlnLysPheValasp----- 71
Db 831 AAATCAACAGATGCTGAATGTGTGACAAAGATGCTGAGAGAAAGATTCAGAAAGCT 890
QY 72 -----GluasnPheThrGluAsnTyrIleuThrAspCysGluGlyLysAs 86
Db 891 TAGTGAATAGAGAGTATGAGAAATCTGAAATGAATTAACAAGTGT---GSTRAGAC 947
QY 86 PALAGLYASnAlaIagIyThrcLysaspGlnSerAspGluValaspGluasp 103
Db 948 TTCAGGTGATGACATGGAAGTGAAGATGATGAAGAGAG---GATGAAGAT 996

RESULT 15
AAK84453 standard; DNA; 1116 BP.
XX AAK84453;
XX
XX 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:39265.
XX
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cycostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-019874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220964.
PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225457.
PR 14-AUG-2000; 2000US-0225738.
PR 14-AUG-2000; 2000US-0225739.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0228287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234323.
PR 21-SEP-2000; 2000US-0234374.
PR 25-SEP-2000; 2000US-0234397.
PR 25-SEP-2000; 2000US-0234598.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 29-SEP-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241121.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.


```
Search completed: February 25, 2003, 02:38:20
Job time : 233 secs
```

PR	08-NOV-2000	2000US-0246477.
PR	08-NOV-2000	2000US-0246478.
PR	08-NOV-2000	2000US-0246523.
PR	08-NOV-2000	2000US-0246524.
PR	08-NOV-2000	2000US-0246525.
PR	08-NOV-2000	2000US-0246526.
PR	08-NOV-2000	2000US-0246527.
PR	08-NOV-2000	2000US-0246528.
PR	08-NOV-2000	2000US-0246532.
PR	08-NOV-2000	2000US-0246569.
PR	08-NOV-2000	2000US-0246611.
PR	08-NOV-2000	2000US-0246613.
PR	17-NOV-2000	2000US-0249207.
PR	17-NOV-2000	2000US-0249208.
PR	17-NOV-2000	2000US-0249209.
PR	17-NOV-2000	2000US-0249210.
PR	17-NOV-2000	2000US-0249211.
PR	17-NOV-2000	2000US-0249212.
PR	17-NOV-2000	2000US-0249213.
PR	17-NOV-2000	2000US-0249214.
PR	17-NOV-2000	2000US-0249215.
PR	17-NOV-2000	2000US-0249216.
PR	17-NOV-2000	2000US-0249217.
PR	17-NOV-2000	2000US-0249218.
PR	17-NOV-2000	2000US-0249244.
PR	17-NOV-2000	2000US-0249245.
PR	17-NOV-2000	2000US-0249264.
PR	17-NOV-2000	2000US-0249265.
PR	17-NOV-2000	2000US-0249297.
PR	17-NOV-2000	2000US-0249299.
PR	01-DEC-2000	2000US-0249300.
PR	01-DEC-2000	2000US-0250160.
PR	01-DEC-2000	2000US-0250391.
PR	05-DEC-2000	2000US-0251030.
PR	05-DEC-2000	2000US-0251988.
PR	06-DEC-2000	2000US-0256719.
PR	08-DEC-2000	2000US-0251479.
PR	08-DEC-2000	2000US-0251856.
PR	08-DEC-2000	2000US-0251866.
PR	08-DEC-2000	2000US-0251989.
PR	11-DEC-2000	2000US-0251990.
PR	11-DEC-2000	2000US-0254097.
PR	05-JAN-2001	2001US-0259678.
PA	(HUMA-) HUMAN GENOME SCI INC.	
PI	Rosen CA, Barash SC, Ruben SM;	
DR	WPI; 2001-483426/52.	
XX		
PT	Nucleic acids encoding human immune/hematopoietic antigen polypeptides,	
XX	useful for preventing, diagnosing and/or treating cancers and	
XX	metastasis -	
PS	Disclosure; SEQ ID NO 39265; 3071pp + Sequence Listing; English.	
XX		
CC	AAK4951 to AAK4702 encode the human immune/hematopoietic antigen (I)	
CC	amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic	
CC	activity, and can be used in gene therapy and vaccine production. (I)	
CC	proteins and polynucleotides may be used in the prevention, diagnosis and	
CC	treatment of diseases associated with inappropriate (I) expression. For	
CC	example, they may be used to treat disorders associated with decreased	
CC	expression by rectifying mutations or deletions in a patient's genome	
CC	that affect the activity of (I) by expressing inactive proteins or to	
CC	supplement the patients own production of (I). Additionally, (I)	
CC	polynucleotides may be used to produce the secreted (I). (I)	
CC	the nucleic acids into a host cell and culturing the cell to express the	
CC	protein. (I) proteins and polynucleotides may be used to prevent,	
CC	diagnose and treat immune/hematopoietic-related diseases, especially	
CC	cancers and cancer metastases of hematopoietic-derived cells. AAK64703	
CC	to AAK87694 represent human immune/hematopoietic antigen genomic	

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 25, 2003, 02:33:25 ; Search time 48 Seconds

(without alignments)
658.077 Million cell updates/sec

Title: US-09-936-737A-2

Perfect score: 573
Sequence: 1 EREDDCWFANRKYIDFDK.....GKDNAGNAGTDESDVEDD 103

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL=framed+ p2n.model -DEV=xlh
-O=/cgn2.1/USPTO.spool/US09936737/rnat_14022003_100733_2308/app-query.fasta-1.263
-DB=Issued Patents_NA -QPM=fastap -SUFFIX=tnl -MIMATCH=0.1 -LOOPL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsun62 -TRANS=human4.0.cdi
-LIST=43 -DOCALLIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15
-MODE=LOCAL -OUTEWT=plc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09936737 @CGN.1.1.25 @unat.14022003.100733.2308 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -LONGLOG -DEV_TIMEDEXT=120
-WARN_TIMEDEXT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents_NA.*
1: /cgn2.6/prodata/1/ina/5A.COMB.seq.*
2: /cgn2.6/prodata/1/ina/5B.COMB.seq.*
3: /cgn2.6/prodata/1/ina/6A.COMB.seq.*
4: /cgn2.6/prodata/1/ina/6B.COMB.seq.*
5: /cgn2.6/prodata/1/ina/PCUT5.COMB.seq.*
6: /cgn2.6/prodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	81.5	14.2	2255	2	US-08-741-134-1
2	80.5	14.0	378	1	US-08-044-547-4
3	80.5	13.2	441	1	US-08-044-547-2
4	75.5	13.2	1402	3	US-08-258-287B-56
5	75.5	13.2	1402	3	US-08-368-704C-54
6	69.5	12.1	2023	3	US-08-961-083-199
7	69.5	12.1	32768	4	US-08-961-527-71
8	68	11.9	1540	4	US-09-149-476-151
9	68	11.9	1676	4	US-08-484-105-21
10	68	11.9	1676	4	US-08-484-105-21
11	68	11.9	2196	4	US-09-149-476-289
12	67.5	11.8	999	6	5514567-5

Result	Sequence	Alignment
13	67.5 11.8 4379 1	US-08-592-214A-17
14	67.5 11.8 4379 3	US-09-149-976-17
15	67 11.7 2143 4	US-09-069-023-21
16	67 11.7 3110 4	US-08-936-165A-257
17	66.5 11.6 261 4	US-09-134-001C-1149
18	66.5 11.6 7295 2	US-08-487-826B-15
19	65.5 11.4 1179 4	US-09-134-001C-251
20	65.5 11.4 1647 4	US-09-472-971-5
21	65.5 11.4 2560 4	US-08-961-527-114
22	65.5 11.4 4107 2	US-08-813-990-24
23	65.5 11.4 4524 4	US-09-472-971-7
24	64 11.2 2276 4	US-09-172-841-56
25	63.5 11.1 2735 5	US-08-920-095-1
26	63.5 11.1 2735 5	PCR-US96-05800-1
27	63.5 11.1 11309 4	US-08-961-527-108
28	63 11.0 5285 2	US-08-609-049A-29
29	63 11.0 5285 4	US-09-170-996-29
30	62.5 10.9 10091 3	US-09-058-489-34
31	62 10.8 28804 2	US-08-592-874-1
32	62 10.8 28804 3	US-09-096-942-2
33	62 10.8 28804 3	US-09-096-942-2
34	61.5 10.7 954 4	US-08-462-625-37
35	61.5 10.7 954 4	US-08-462-625-37
36	61.5 10.7 988 4	US-08-098-327E-34
37	61.5 10.7 988 4	US-08-462-625-34
38	61.5 10.7 1251 2	US-09-211-930-2
39	61.5 10.7 1251 4	US-09-340-993-2
40	61.5 10.7 1251 4	US-09-468-442-2
41	61.5 10.7 1345 2	US-08-702-153-3
42	61.5 10.7 1353 3	US-09-211-930-8
43	61.5 10.7 1353 3	US-09-340-993-8
44	61.5 10.7 1353 3	US-09-468-442-8
45	61.5 10.7 2191 4	US-08-875-082-1

ALIGNMENTS

RESULT 1
US-08-741-134-1
Sequence 1, Application US/08741134
Patent No. 5861498
GENERAL INFORMATION:
APPLICANT: Litwack, Gerald
APPLICANT: Alnemri, Emad S.
TITLE OF INVENTION: Fernandez-Alnemri, Teresa
TITLE OF INVENTION: IMMUNOPHILIN FKBP46 AND COMPOSITIONS FOR MAKING
NUMBER OF INVENTIONS: AND
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5861498rls
STREET: One Liberty Place - 46th floor
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 3.11
SOFTWARE: Wordperfect for Windows 6.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/741.134
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/007,163
FILING DATE: 01-NOV-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Deluca, Mark
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: TJU-2090

```

TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-568-3100
TELEFAX: 215-568-3439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2255 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: not relevant
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 160..1398
US-08-741-134-1

Alignment Scores:
Pred. No.: 0.187 Length: 2255
Score: 81.50 Matches: 33
Percent Similarity: 43.12% Conservative: 14
Best Local Similarity: 30.28% Mismatches: 34
Query Match: 14.22% Indels: 28
DB: Gaps: 5

US-09-936-737a-2 (1-103) x US-08-741-134-1 (1-2255)
QY 1 GlnGluArgLysCysTyrPheThrAlaAsnArgLysTyrThrAspPheAspLys 20
   |||||
DB 525 GAGGAAGCTCGAAGATGCAATGATGCCACGCTAACAAAGGCCACCTGACAGAA 584
QY 20 sSerPheLysLysSer-----SerAspLeuAspGluCysLysTyrCys 35
   |||||
DB 585 AGCTGGCAAGAACAGTGCACGACGACGAGAAAGTGAATGATGATGATGATGAT 636
QY 35 sPheLysThrLysTyrCysTyrLysLeuPheGluAspThrValAsnLysGluCysTyr 55
   |||||
DB 637 -----GAGACCACTTCAAG----- 654
QY 55 AsnValAlaAspGluGluGluLeuAspGluGluLysPheValValAspGluAsnPh 75
   |||||
DB 655 -----TTCTTCGACGCTGAGATATAGACACGATGAA-----AATGATGATCATTC 704
QY 75 rGluAsnTyrLeuThrAspCysGluGluLysAspAlaGluAsnAlaLysThrGly 95
   |||||
DB 705 AATGAAC-----ACATCAGCTGAAGGAGATGACAGATGAAGAGATGATGAAGA 758
QY 95 pGluSerAspGluValAspGluAsp 103
   |||||
DB 759 CGAAGAGATGAGAAAGATGATGAT 783

RESULT 2
US-08-044-547-4
; Sequence 4, Application US/08044547
; Patent No. 5324715
GENERAL INFORMATION:
APPLICANT: Connolly, Thomas M.
APPLICANT: Keller, Paul M.
TITLE OF INVENTION: Protein for Inhibiting
TITLE OF INVENTION: Collagen-stimulated Platelet Aggregation
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merck & Co., Inc.
STREET: P.O. Box 2000
CITY: Rahway
STATE: New Jersey
COUNTRY: US
ZIP: 07065
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/044,547

```

```

FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/594,917
FILING DATE: 09-OCT-1990
ATTORNEY/AGENT INFORMATION:
NAME: Parr, Richard S.
REGISTRATION NUMBER: 32,586
REFERENCE/DOCKET NUMBER: 18053
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 594-4958
TELEFAX: (908) 594-4720
TELEX: 138825
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 378 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
US-08-044-547-4

Alignment Scores:
Pred. No.: 0.0232 Length: 378
Score: 80.50 Matches: 18
Percent Similarity: 50.91% Conservative: 10
Best Local Similarity: 32.73% Mismatches: 24
Query Match: 14.05% Indels: 3
DB: Gaps: 1

US-09-936-737a-2 (1-103) x US-08-044-547-4 (1-378)
QY 5 AspCysTyrPheThrPheTyrAlaAsnArgLysTyrThrAspPheAspLysSerPheLys 24
   |||||
DB 133 GATTCGTGCTGAAGAGCGCTGCGTGAAGCGCTGACAAATCTGTGACGAAGACTGA 192
QY 25 SerSerAspLeuAspGluCysLysTyrCysPheLysThrGluTyr-----Cys 41
   |||||
DB 193 TTCACGCGGCTGATGAAGACGAAGAGATGTCGAAGATTCGCCGTGCAACATCTGCG 252
QY 42 TyrIleValPheGluAspThrValAsnLysGluCysTyrTyrAsn 56
   |||||
DB 253 TACATTCGCAATCAACACAGAAACCAACGATGCTATAGAAAT 297

RESULT 3
US-08-044-547-2
; Sequence 2, Application US/08044547
; Patent No. 5324715
GENERAL INFORMATION:
APPLICANT: Connolly, Thomas M.
APPLICANT: Keller, Paul M.
TITLE OF INVENTION: Protein for Inhibiting
TITLE OF INVENTION: Collagen-stimulated Platelet Aggregation
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merck & Co., Inc.
STREET: P.O. Box 2000
CITY: Rahway
STATE: New Jersey
COUNTRY: US
ZIP: 07065
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/044,547
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/594,917
FILING DATE: 09-OCT-1990

```

ATTORNEY/AGENT INFORMATION:
NAME: Parr, Richard S.
REGISTRATION NUMBER: 32,586
REFERENCE/DOCKET NUMBER: 18053
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 594-4958
TELEFAX: (908) 594-4720
TELEX: 138825
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 441 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
US-08-044-547-2

Alignment Scores:
Pred. No.: 0.0285 Length: 441
Score: 80.50 Matches: 18
Percent Similarity: 50.91% Conservative: 10
Best Local Similarity: 32.73% Mismatches: 24
Query Match: 14.05% Indels: 3
DB: 1 Gaps: 1

US-09-936-737A-2 (1-103) x US-08-044-547-2 (1-441)
QY 5 AspCysTrpThrPheTyraIaenATgLySTyTThrAspPheAspLysSerPheLysLys 24
DB 196 GATTGCTGCTTAACGGCGCTGCTGAAGCTGCTGACATCTGTGACGAAGACTGAA 255
QY 25 SerSerAspLeuAspLysLysLysLysLysLysPheLysThrGluTyr-----Cys 41
DB 256 TTCACACGGCTCGATGATGAAGAAAGATGTCGAAGATCGCGCTGGAACCATCTGC 315
QY 42 TyrlleValPheGluAspThrValAsnLysGluCysTyTyrTrAsn 56
DB 316 TACATCTGCAATCAACACAGAACCAACGATCTATAGAAAT 360

RESULT 4
US-08-258-287B-56
Sequence 56, Application US/08258287B
Patent No. 6083735
GENERAL INFORMATION:
APPLICANT: Yuan, Junying
APPLICANT: Miura, Masayuki
TITLE OF INVENTION: Programmed Cell Death Genes and Proteins
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/258,287B
FILING DATE: 10-JUN-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/080,850
FILING DATE: 24-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Bugalsky, Lawrence B.
REGISTRATION NUMBER: 35,086
REFERENCE/DOCKET NUMBER: 0609.3920001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600

TELEFAX: (202) 371-2540
TELEX: 248636 SSK
INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 1402 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: CDS
LOCATION: 45..1301
US-08-258-287B-56

Alignment Scores:
Pred. No.: 0.635 Length: 1402
Score: 75.50 Matches: 26
Percent Similarity: 54.79% Conservative: 14
Best Local Similarity: 35.62% Mismatches: 14
Query Match: 13.18% Indels: 19
DB: 3 Gaps: 6

US-09-936-737A-2 (1-103) x US-08-258-287B-56 (1-1402)
QY 44 ValPheGluAspThrValAsnLysGluCysTyTyrTrAsnValAspGluGluLeu 63
DB 120 GTTTTGATGATGCTGCTGCGAGAG-----AATGTTTAAATGAGATGATGTTA 167
QY 64 -----AspGluGluLysPheValValAsp-----GluAsnPheThrGluAsn 77
DB 168 CTAAATAATAGGGGAAGGCGAGTTTCATCTGACAGCGCTGAAGAAATCTGTTGAGAAC 227
QY 78 Tyrlleu-----ThrAspCysGluGluLys---AspAlaGlyAsnAlaIaIaGlyThIgly 94
DB 228 TTCCTTAGAGAAAACAGACATGCGAGAAAATATTGCTGCGCCATTTGCCAATTTCCAG 287
QY 95 AspGlu-----SerAspGluValAsp 101
DB 288 GAACAGCTGATGTTACAAATTTCTAATGATGATGATGAT 326

RESULT 5
US-08-368-704C-54
Sequence 54, Application US/08368704C
Patent No. 6087160
GENERAL INFORMATION:
APPLICANT: Yuan, Junying
APPLICANT: Miura, Masayuki
TITLE OF INVENTION: Programmed Cell Death Genes and Proteins
NUMBER OF SEQUENCES: 95
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/368,704C
FILING DATE: 4-JAN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/258,287
FILING DATE: 10-JUN-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/080,850
FILING DATE: 24-JUN-1993
ATTORNEY/AGENT INFORMATION:

NAME: Bugalsky, Lawrence B.
 REGISTRATION NUMBER: 35,086
 REFERENCE/DOCKET NUMBER: 0609.3920002
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 371-2600
 TELEFAX: (202) 371-2540
 TELETYPE: 248636 SSK
 INFORMATION FOR SEQ ID NO: 54:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1402 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: both
 TOPOLOGY: both
 MOLECULE TYPE: CDNA
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 45..1301
 US-08-368-704C-54

Alignment Scores:
 Pred. No.: 0.635 Length: 1402
 Score: 75.50 Matches: 26
 Percent Similarity: 54.79% Conservative: 14
 Best Local Similarity: 35.62% Mismatches: 14
 Query Match: 13.18% Indels: 19
 DB: Gaps: 6

US-09-936-737a-2 (1-103) x US-08-368-704C-54 (1-1402)

QY 44 ValPheGluAspThrValAsnGlyCysTyrTyrAspValValAspGluGluLeu 63
 |||||:||||| ||||| |||||:|||||:|||||
 Db 120 GTTTTGATGACCTGGGAGAG-----AATGTTTAATGAGATAGTTA 167
 QY 64 -----AspGlnGluyspHeValValAsp-----GluAspHeThrGluAsn 77
 |||||:||||| |||||:|||||
 Db 168 CTCAAATAGGGAAGTGGAGTTTCATCCGACAGCTGAGATCTGGTGAAGAC 227
 QY 78 TyrLeu-----ThrAspCysGluGlyLys-----AspAlaGlyAsnAlaGlyThrGly 94
 |||||:||||| |||||:|||||:|||||
 Db 228 TTCTTAGAGAAACAGACATGCGAGAAATATTTGCTGGCCACTTCCCAATTCGCCAG 287
 QY 95 AspGlu-----SerAspGluValAsp 101
 |||||:||||| |||||
 Db 288 GAACAGCTGAGTTACAAATTTCTAATGATGAGCATGAT 326

RESULT 6

US-08-961-083-199

; Sequence 199, Application US/08961083

; Patent No. 6159469

; GENERAL INFORMATION:

APPLICANT: Choi et. al.

TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines

NUMBER OF SEQUENCES: 452

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/961.083

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Brookes, A. Anders
 REGISTRATION NUMBER: 36,373
 REFERENCE/DOCKET NUMBER: PB340P2
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (301) 309-8504
 TELEFAX: (301) 309-8512
 INFORMATION FOR SEQ ID NO: 199:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2023 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 US-08-961-083-199

Alignment Scores:
 Pred. No.: 6.68 Length: 2023
 Score: 69.50 Matches: 25
 Percent Similarity: 39.62% Conservative: 17
 Best Local Similarity: 23.58% Mismatches: 35
 Query Match: 12.13% Indels: 29
 DB: Gaps: 5

US-09-936-737a-2 (1-103) x US-08-961-083-199 (1-2023)

QY 7 TrpThrPheTyrAlaAsnArgLysTyrThrAspPheAspLysSerPheLysSerSer 26
 |||||:||||| ||||| |||||:|||||
 Db 965 TGGATTAGCAATTAATATAGCGCTTAGAGAGATTGAC-----TCA 1006
 QY 27 AspLeuAspGluCysLysLysThr-----CysPheLysThrGluTyrCys 41
 |||||:||||| ||||| |||||:|||||
 Db 1007 GAAATATATTAATGTCGGTAACTCTGACCAATTTGATGTATTAATTCGATATAGT 1066
 QY 42 TyrIleValPheGluAspThrValAsnGlyCysTyrTyrAsnValValAspGlyGlu 61
 |||||:||||| ||||| |||||:|||||
 Db 1067 TACACAGCTCTT-----TTACCTATTTCAATGCTGATTCGTG 1105
 QY 62 GluLeuAspGlnGluLysPhe-----ValValAspGluAsnPheThrGlu 76
 |||||:||||| ||||| |||||:|||||
 Db 1106 CAGAGACAGAGCCCTTCACTTGACGTGATCTAGTTGTAACGAAAAATCTGCATGAC 1165
 QY 77 AsnTyrLeuThrAspCysGluGly-----LysAspAlaGlyAsnAla 90
 |||||:||||| ||||| |||||:|||||
 Db 1166 TTGTTCTCTACAGACTTACAGATTAATCCTTGGCTGCTGTAGAGATTTGGGGCAGA 1225
 QY 91 AlaGlyThrGlyAspGlu 96
 |||||:||||| |||||
 Db 1226 GCTTATTTTGGTCAGAA 1243

RESULT 7

US-08-961-527-71

; Sequence 71, Application US/08961527

; Patent No. 6420135

; GENERAL INFORMATION:

APPLICANT: Charles Kunsch

TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences

NUMBER OF SEQUENCES: 391

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/961.527

FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

```

; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brooks, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 71:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 32768 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
US-08-961-527-71

Alignment Scores:
Pred. No.: 280 Length: 32768
Score: 69.50 Matches: 25
Percent Similarity: 39.62% Conservative: 17
Best Local Similarity: 23.58% Mismatches: 35
Query Match: 12.13% Indels: 29
DB: 4 Gaps: 5

US-09-936-737A-2 (1-103) x US-08-961-527-71 (1-32768)

QY 7 TrpThrPheTyrAlaAsnArgLysTyrThrAspPheAspLysSerPheLysSerSer 26
Db 7201 TGGATTAGCAATTAATTAAGCCGCTTAGAAGACTTTGAC-----TCA 7242
QY 27 AspLeuAspGluCysLysLysThr-----CysPheLysThrGluTyrCys 41
Db 7243 GAAATTTTATTAATGTCGCGTAACCTCTGACAAATTTCAATGTTATTAACCGATATTACT 7302
QY 42 TyrIleValPheGluAspThrValAsnLysGluCysTyrTyrAsnValAlaSpGlu 61
Db 7303 TACCACTCTT-----TTACGCTATTTCATACCTGATTTGCGT 7341
QY 62 GluLeuAspGluLysPhe-----ValValAspGluAsnPhethrGlu 76
Db 7342 CAGAGACACAGAGCCCTCTGACTGTGACTGTGATCTGATGTAACGAAATCTGATGAC 7401
QY 77 AsnTyrIleuThrAspCysGluGly-----LysAspAlaGluAsnAla 90
Db 7402 TTGTTGCTACAGACTTACAGATTATCTTGGCTGCTGTAGAGATTTGGGGCAGA 7461
QY 91 AlaGlyThrGluAspGlu 96
Db 7462 GCTTATTTTGTCAAGAA 7479

RESULT 8
US-09-149-476-151/c
; Sequence 151, Application US/09149476
; Patent No. 6420526
; GENERAL INFORMATION:
; APPLICANT: Rosen et al
; TITLE OF INVENTION: 186 Human Secreted proteins
; FILE REFERENCE: P2002P1
; CURRENT APPLICATION NUMBER: US/09/149,476
; EARLIER FILING DATE: 1998-09-08
; EARLIER APPLICATION NUMBER: PCT/US98/04493
; EARLIER FILING DATE: 1998-03-06
; EARLIER APPLICATION NUMBER: 60/040,162
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,333
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/038,621
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,626
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,334
; EARLIER FILING DATE: 1997-03-07
```

```

; EARLIER APPLICATION NUMBER: 60/040,336
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,163
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/047,600
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,615
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,597
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,502
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,633
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,583
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,617
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,618
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,503
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,592
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,581
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,584
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,500
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,587
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,492
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,598
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,613
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,582
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,596
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,612
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,632
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,601
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,580
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,568
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,314
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,569
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,311
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,671
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,674
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,669
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,312
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,313
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,672
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,315
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/048,974
```

```

; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/056,886
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,877
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,889
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,893
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,630
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,878
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,662
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,872
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,882
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,637
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,903
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,888
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,879
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,880
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,894
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,911
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,636
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,874
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,910
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,864
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,631
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,845
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,892
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/057,761
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/047,595
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,559
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,588
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,585
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,586
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,590
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,594
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,589
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,593
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,614
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,578
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,576
; EARLIER FILING DATE: 1997-04-11

```

```

; EARLIER APPLICATION NUMBER: 60/047,501
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,670
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/056,632
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,664
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,876
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,881
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,909
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,875
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,862
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,887
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,908
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/048,964
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/057,650
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: 60/056,884
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/057,669
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: 60/049,610
; EARLIER FILING DATE: 1997-06-13
; EARLIER APPLICATION NUMBER: 60/061,060
; EARLIER FILING DATE: 1997-10-02

```

Alignment Scores:

```

Pred. No.: 7.37 Length: 1540
Score: 68.00 Matches: 27
Percent Similarity: 40.00% Conservative: 15
Best Local Similarity: 25.71% Mismatches: 41
Query Match: 11.87% Indels: 22
DB: 4 Gaps: 5

```

US-09-936-737a-2 (1-103) x US-09-149-476-151 (1-1540)

```

QY 9 PheTYAlaAsnArgLysTYR-----ThsAspPheasp 19
||||:|||||
Db 1111 TTTGGGCAATGGAATATCATATGCGCCATTGCCACAAGAGCAATCAAT 1052
||||:|||||
QY 20 LysSerPheLysLysSerAspLeuAspGluCysLysLysThCysPheLys----- 37
||||:|||||
Db 1051 ACGAGCAAGCAAAAGCTTCAGATTGGA--GCCAGATGACCATGATGAAATGCAGA 995
||||:|||||
QY 38 -----ThrgLysTYR-----TyrlleValPhegluAspThr 48
||||:|||||
Db 994 AGATCAAGAAAGATGATCATGAGATTCTGTACCCATTTATATAGCTCTTCAATCA 935
||||:|||||
QY 49 ValAsnLysglucysTYrTYrAsnValValAspGluGluLeuAspLingLysPhe 68
||||:|||||
Db 934 ATTACATTTTCTTCC--ACCACTGTGTACAGTTAAATCAAGAGCAACCAAGCCATT 878
||||:|||||
QY 69 ValValAspLysPheLysPheThrGluAsnTYrLysThsAspCysGluGluLysAspAlaGly 88
||||:|||||
Db 877 CTACACAGCAATGACGCTGCACAAACAGCAGACGTGACAGATATATGCGCACAGATTCCAGA 818
||||:|||||
QY 89 AsnAlaAlaGlyThr 93
||||:|||||
Db 817 CTGTGTCCGGGACA 803
||||:|||||

```

```

RESULT 9
US-08-484-105-21
; Sequence 21 Application us/08484105
; Patent No. 5589341

```



```

: GENERAL INFORMATION:
: APPLICANT: STILLMAN, Bruce
: APPLICANT: BELL, Stephen P
: APPLICANT: KOBAYASHI, Ryuji
: APPLICANT: RINE, Jasper
: APPLICANT: FOSS, Margit
: APPLICANT: MCNALLY, Francis J
: APPLICANT: LAURENSEN, Patricia
: APPLICANT: HERSKOWITZ, Ira
: APPLICANT: LI, Joachim J
: APPLICANT: GAVIN, Kimberly
: TITLE OF INVENTION: ORIGIN OF REPLICATION COMPLEX GENES
: NUMBER OF SEQUENCES: 24
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
: STREET: 4 Embarcadero Center, Suite 3400
: CITY: San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94111-4187
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patent In Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/484,105
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Osman Ph.D., Richard Aron
: REGISTRATION NUMBER: 36,627
: REFERENCE/DOCKET NUMBER: A-59032/DJB/RAO
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 494-8700
: TELEFAX: (415) 494-8771
: TELETYPE: 910 277299
: INFORMATION FOR SEQ ID NO: 21:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1676 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: CDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 13..1302
: US-08-484-105-21
:
: Alignment Scores:
: Pred. No.: 8.26 Length: 1676
: Score: 68.00 Matches: 29
: Percent Similarity: 45.54% Conservative: 22
: Best Local Similarity: 25.89% Mismatches: 24
: Query Match: 11.87% Indels: 37
: Gaps: 7
:
: US-09-936-737a-2 (1-103) x US-08-484-105-21 (1-1676)
:
: QY 20 LysSerPheLysLysSerSerAspLeuAspGluCysLysLysThrCysPheLys----- 37
: |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
: Db 24 AAAAATTTTGAACGACGACGCTCCAGCCAGTCCGCGCTCTGTGAAAAAATCGAC 83
:
: QY 38 -----ThnGluTyrCysTyrIleValPheGluAspThrValAsn 50
: |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
: Db 84 TCCGAAAAAGAGATCCAGACAGAA-----AAAGACGAATGGAAA 125
:
: QY 51 LysGluCysTyr-TyrAsnValValAspGly-----GluGluLeuAspGln----- 65
: :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
: Db 126 AGAGAAATGCTTCTAGAAATTTGCAATCAATTTAGAAAGAGATTGGAAACAATCGGCTT 185
:
: QY 66 -----GluLysPheValValAspGluAsnPheThnGluAsnTyrIleThrAspGly 83
: |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

Db 186 CGAGATGAACCTGTATCAATGCGTCAATCAGCAATCGAAATTACTTT-----ATGCA 239
:
: QY 83 uGlyLysAspPala-----GlyAsnAlaAla 92
: :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
: Db 240 AGGAAAAATCGCGCTCAGAACGAATGAATATGCGAAATCCGCTGTGGAAGACGTCTGG 299
:
: QY 92 yThrGlyAspGluSerAspGluValAspGluAsp 103
: |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
: Db 300 AAATGGAAAT--ACTGAAGAAATTGAGGAAGAC 330
:
: RESULT 10
: US-08-484-106-21
: Sequence 21, Application US/08484106
: Patent No. 5614518
:
: GENERAL INFORMATION:
: APPLICANT: STILLMAN, Bruce
: APPLICANT: BELL, Stephen P
: APPLICANT: KOBAYASHI, Ryuji
: APPLICANT: RINE, Jasper
: APPLICANT: FOSS, Margit
: APPLICANT: MCNALLY, Francis J
: APPLICANT: LAURENSEN, Patricia
: APPLICANT: HERSKOWITZ, Ira
: APPLICANT: LI, Joachim J
: APPLICANT: GAVIN, Kimberly
: TITLE OF INVENTION: ORIGIN OF REPLICATION COMPLEX GENES
: NUMBER OF SEQUENCES: 24
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
: STREET: 4 Embarcadero Center, Suite 3400
: CITY: San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94111-4187
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patent In Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/484,106
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Osman Ph.D., Richard Aron
: REGISTRATION NUMBER: 36,627
: REFERENCE/DOCKET NUMBER: A-59032/DJB/RAO
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 494-8700
: TELEFAX: (415) 494-8771
: TELETYPE: 910 277299
: INFORMATION FOR SEQ ID NO: 21:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1676 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: CDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 13..1302
: US-08-484-106-21
:
: Alignment Scores:
: Pred. No.: 8.26 Length: 1676
: Score: 68.00 Matches: 29
: Percent Similarity: 45.54% Conservative: 22
: Best Local Similarity: 25.89% Mismatches: 24
: Query Match: 11.87% Indels: 37
: Gaps: 7
:
: US-09-936-737a-2 (1-103) x US-08-484-106-21 (1-1676)

```

```
QY 20 LysSerPheLysLysSerSerAspLeuAspGluCysLysLysThrCysPheLys----- 37
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 24 AAAAATTTGAACGAGCACTGTCCAGCCGCGCCCTCTCTGAAAAAATCGAC 83
QY 38 -----ThGluTyrCysTyrIleValPheGluAspThrValAsn 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 84 TCCAGAAAAAGAGATCCAGACAGAA-----AAAGCAATGAGAA 125
QY 51 LysGluCysTyr-TyrAsnValValAspGly-----GluGluLeuAspGln----- 65
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 126 AGAGAACTCTCTAGAAATTTGCAATCAATTTAGAGAGAAATTTGAGAACACTGGGCTT 185
QY 66 -----GluLysPheValValAspGluAsnPheThrGluAsnTyrLeuThrAspCysG1 83
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 186 CGAGATCAAACTGTATCAATGCGCTCAATCGCAATCGAAATATCTTT-----ATGCA 239
QY 83 uGlyLysAspAla-----GlyAsnAlaAlaG1 92
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 240 AGGAAATCGCGCTCAGAACGAAATGAATATCGAAATCCCGCTCGTGAAGACGTGCTGG 299
QY 92 yThrGlyAspGluSerAspGluValAspGluAsp 103
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 300 AATGCAAT--ACTGAAGAAATTTGAGAGAC 330

RESULT 11
US-09-149-476-289/C
Sequence 289, Application US/09149476
Patent No. 6420526
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: 186 Human Secreted proteins
FILE REFERENCE: P2002P1
CURRENT APPLICATION NUMBER: US/09/149, 476
EARLIER FILING DATE: 1998-09-08
EARLIER APPLICATION NUMBER: PCT/US98/04493
EARLIER FILING DATE: 1998-03-06
EARLIER APPLICATION NUMBER: 60/040,162
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,333
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/038,621
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,626
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,334
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,336
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,163
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/047,600
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,615
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,597
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,502
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,633
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,583
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,617
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,618
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,503
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,592
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,581
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,584

EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,500
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,587
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,492
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,598
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,613
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,582
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,596
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,612
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,632
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,601
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/043,580
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,568
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,314
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,569
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,311
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,671
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,674
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,669
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,312
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,313
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,672
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,315
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/048,974
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/056,886
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,877
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,889
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,893
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,630
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,878
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,662
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,872
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,882
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,637
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,903
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,888
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,879
EARLIER FILING DATE: 1997-08-22
```

[REDACTED]

Downloaded from <http://ajph.org/> on November 10, 2014

Oy	16	ThrsAppPhasplySerSperhelysSersSerasp-----LeuaspGluCySylslys 33
		::: ::::
Db	1119	ACGATTTGTTCACGACAGATTTCAGAAAGAGACGATCGACAGATTAATTGTAGTAT 1060
Oy	34	ThCySperhelysrthGluYrCySyrTLLeValPhegluaSperthValasnlYsgluys 53
		::: ::: :::
Db	1059	TCAAAATTTCAAACCCAAATAATGATTTTGTGTC---AAGATTTTTCATGTTGTGTGT 1003
Oy	54	TyTyAsnValaValasplYsgluGluLeuaspGln 65
		:::
Db	1002	TCTTATTAGTGTGTAGATGATCAGACAAAATACAG 967
		RESULT 14
		US-09-149-976-17/C
		: Sequence 17, Application US/09149976
		: Patent No. 6127123
		GENERAL INFORMATION:
		APPLICANT: Yanofsky, Martin F.
		TITLE OF INVENTION: Cauliflower Floral Meristem Identity
		: TITLE OF INVENTION: Genes and Methods of Using Same
		NUMBER OF SEQUENCES: 33
		CORRESPONDENCE ADDRESS:
		ADDRESSEE: Campbell & Flores LLP
		STREET: 4370 La Jolla Village Drive, Suite 700
		CITY: San Diego
		STATE: California
		COUNTRY: United States
		ZIP: 92122
		COMPUTER READABLE FORM:
		MEDIUM TYPE: Floppy disk
		COMPUTER: IBM PC compatible
		OPERATING SYSTEM: PC-DOS/MS-DOS
		SOFTWARE: Patentin Release #1.0, Version #1.25
		CURRENT APPLICATION DATA:
		APPLICATION NUMBER: US/09/149,976
		FILING DATE: 09-SEP-1998
		CLASSIFICATION: 435
		PRIOR APPLICATION DATA:
		APPLICATION NUMBER: US 08/592,214
		FILING DATE: 26-JAN-1996
		ATTORNEY/AGENT INFORMATION:
		NAME: Campbell, Cathryn A.
		REGISTRATION NUMBER: 31,815
		REFERENCE/DOCKET NUMBER: P-UD 3291
		TELECOMMUNICATION INFORMATION:
		TELEPHONE: (619) 535-9001
		TELEFAX: (619) 535-8949
		INFORMATION FOR SEQ. ID NO.: 17:
		SEQUENCE CHARACTERISTICS:
		LENGTH: 4379 base pairs
		TYPE: nucleic acid
		STRANDEDNESS: double
		TOPOLOGY: linear
		FEATURE:
		NAME/KEY: unsure
		LOCATION: 2095..2098
		OTHER INFORMATION: /note= "N = one or more
		OTHER INFORMATION: nucleotides."
		FEATURE:
		NAME/KEY: misc_feature
		LOCATION: 1..4379
		OTHER INFORMATION: /note= "sequence = Arabidopsis
		OTHER INFORMATION: thaliana API gene"
		US-09-149-976-17
		Alignment Scores:
		Pred. No.: 35 Length: 4379
		Score: 67.50 Matches: 19
		Percent Similarity: 53.85% Conservative: 9
		Best Local Similarity: 36.54% Mismatches: 21
		Query Match: 11.78% Indels: 3
		Query: 3 Gaps: 2

US-09-936-737A-2 (1-103) x US-09-149-976-17 (1-4379)

QY 16 ThrAspPheAspLysSerPheLysSerSerAsp-----LeuAspGluCysLysLys 33
Db 1119 ACGGATTTTACGACGAGATTAAAGAGAGAGAAATCGAGAGATTAATGTAAGTAT 1060
QY 34 ThrCysPheLysThrGluTyrCysTyrIleValPheGluAspThrValAsnLysGluCys 53
Db 1059 TCATAATTTCACAAACCAAAATAGTATTTTGTGC---AAAGATTTTTCATGATGTTGTGT 1003
QY 54 TyrTyrAsnValValAspGluGluLeuAspGln 65
Db 1002 TCTTATTAAGTTGATGATGATCAGCAAAATATCAG 967

RESULT 15

US-09-069-023-21
; Sequence 21, Application US/09069023A
; Patent No. 6348573
; GENERAL INFORMATION:
; APPLICANT: Nunez, Gabriel
; APPLICANT: Inohara, Naohiro
; APPLICANT: Koseki, Takeyoshi
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING APOPTOSIS
; FILE REFERENCE: UM-03333
; CURRENT APPLICATION NUMBER: US/09/069, 023A
; CURRENT FILING DATE: 1998-04-27
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 2143
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-069-023-21

Alignment Scores:

Pred. No.:	15.7	Length:	2143
Score:	67.00	Matches:	32
Percent Similarity:	43.36%	Conservative:	17
Best Local Similarity:	28.32%	Mismatches:	28
Query Match:	11.69%	Indels:	36
DB:	4	Gaps:	9

US-09-936-737A-2 (1-103) x US-09-069-023-21 (1-2143)

QY 14 LysTyrThrAspPheAspLysSerPheLysLys-----SerSerAspLeuAspGlu 30
Db 278 AAAGTCAAGGACGTAACACCAATTAAGAGGCTGTCTCGCGCTCCAGCTCGAGAG 337
QY 31 CysLys---LysThrCysPheLysThrGluTyrCysTyrIleValPheGluAspThrVal 49
Db 338 ATCCGCGAGGAGGCGGAGAGATTCGAGAGTGGACCACTG-----CCCAACATC 391
QY 50 AsnLysGluCysTyrTyrAsnValValAspGluGluLeuAspGlnLysPhe--- 68
Db 392 CACCTGACCTCC-----GACGGCAGCAGATGATGATGAGGATGACTTT 436
QY 69 ---ValValAspGluAsnPheThrGlu----- 76
Db 437 CGACACCTCGACGAGAAAT---ACGGAACTGGTGGCCGCTTTCGCCGAGAAACATTGGATC 493
QY 77 -----AsnTyr-----LeuThrAspCysGluGlyLysAspAlaGly----- 88
Db 494 GATCCACACTGACTGAGTGAATACCACTCCACATGCGACAGGAGGAGAACTGGAAC 553
QY 89 -----AsnAlaAlaGlyThrGlyAspGluSerAsp 98
Db 554 GGAGAGCTGAGCGAGGAGCGGCGGACACACCGAT 592

Search completed: February 25, 2003, 03:23:14
Job time : 53 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 25, 2003, 02:34:15 : Search time 62 Seconds

(without alignments)
933.046 Million cell updates/sec

Title: US-09-936-737A-2

Perfect score: 573
Sequence: 1 EEREDQWTFYANRKYTFDFK.....GMDAGNAGTGESEDEVED 103

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 442118 seqs, 280819700 residues

Total number of hits satisfying chosen parameters: 884236

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=framer-p2n.model -DEV=rlh
-Q=/cgn2_1/USPTO.spool/US0936737/runt_14022003_100733_2323/app_query.fasta_1.263
-DB=Published.Applications.NA -QMT=fastap -SUFFIX=trpb -MIMATCH=0.1
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blousum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=1 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MALEN=200000000 -USER=US0936737 -RCGN_1_1_33_etunat_14022003_100733_2323
-NCPU=6 -ICPU=3 -NO_XLPXY -NO_MMP -LARGEQUERY -NEG SCORE=0 -WAIT -LONGLOG
-DEV TIMEOUT=120 -MARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -XGAPOP=6
-XGAPEXT=7 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Published.Applications.NA:

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	71.5	12.5	1092	10	US-09-815-242-9446
2	70.5	12.3	543	9	US-09-796-692-9477
3	70	12.2	540	9	US-10-040-739-838
4	70	12.2	3504	10	US-09-815-242-7425

5	70	12.2	197997	10	US-09-822-246-3	Sequence 3, Appl1
6	69.5	12.1	2023	10	US-09-765-272-199	Sequence 199, App
7	68	11.9	1861	10	US-09-729-674-129	Sequence 129, App
8	68	11.9	1988	10	US-09-864-761-4740	Sequence 4740, Ap
9	68	11.9	5153	10	US-09-727-384-3	Sequence 3, Appl1
10	67.5	11.8	914	9	US-09-798-889-29	Sequence 29, Appl
11	67.5	11.8	2664	10	US-09-834-975-897	Sequence 897, App
12	67	11.7	763	10	US-09-910-943-179	Sequence 179, App
13	67	11.7	1430	9	US-09-759-1308-61	Sequence 61, Appl
14	67	11.7	2330	9	US-08-759-1308-51	Sequence 51, Appl
15	67	11.7	3110	10	US-09-939-980-237	Sequence 237, App
16	66.5	11.6	3984	9	US-10-149-110-2	Sequence 2, Appl1
17	66.5	11.6	7450	9	US-10-149-110-10	Sequence 10, Appl
18	66.5	11.6	8370	10	US-09-801-574-56	Sequence 56, Appl
19	66.5	11.6	640681	10	US-09-790-988-1	Sequence 1, Appl1
20	66	11.5	634	10	US-09-770-149-663	Sequence 663, App
21	66	11.5	5988	10	US-09-864-864-281	Sequence 281, App
22	66	11.5	640681	10	US-09-790-988-1	Sequence 1, Appl1
23	65.5	11.4	2155	10	US-09-960-428-13	Sequence 13, Appl
24	65	11.3	436	10	US-09-867-701-1512	Sequence 1512, Ap
25	65	11.3	3611	9	US-09-974-298-113	Sequence 113, App
26	64.5	11.3	288	10	US-09-974-300-3913	Sequence 3913, Ap
27	64.5	11.3	469	10	US-09-864-761-6398	Sequence 6398, Ap
28	64.5	11.3	562	10	US-09-954-456-216	Sequence 216, App
29	64.5	11.3	562	10	US-09-954-456-822	Sequence 822, App
30	64.5	11.3	562	10	US-09-954-456-1225	Sequence 1225, Ap
31	64.5	11.3	630	10	US-09-925-297-52	Sequence 52, Appl
32	64.5	11.3	1152	9	US-09-938-842A-2486	Sequence 2486, Ap
33	64	11.2	1777	12	US-10-042-417-51	Sequence 51, Appl
34	64	11.2	2807	10	US-09-940-836A-1	Sequence 1, Appl1
35	64	11.2	2843	9	US-09-822-846-421	Sequence 421, Appl
36	64	11.2	6865	10	US-09-900-038A-3	Sequence 3, Appl1
37	63.5	11.1	411	10	US-09-878-574-4933	Sequence 4933, Ap
38	63.5	11.1	420	10	US-09-878-574-4933	Sequence 1717, Ap
39	63.5	11.1	1629	9	US-09-738-626-1717	Sequence 1, Appl
40	63.5	11.1	1839	10	US-09-981-768A-1	Sequence 180, Ap
41	63	11.0	1839	9	US-09-938-842A-1380	Sequence 7, Appl
42	63	11.0	4166	10	US-09-930-871-7	Sequence 7, Appl1
43	63	11.0	4166	10	US-09-930-871-9	Sequence 17, Appl
44	63	11.0	4179	10	US-09-930-871-17	Sequence 19, Appl
45	63	11.0	4197	10	US-09-930-871-19	

ALIGNMENTS

RESULT 1
US-09-815-242-9446/c
; Sequence 9446, Application US/09815242
; Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl H.
APPLICANT: Walskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27

US-09-936-737A-2 (1-103) X US-09-815-242-9446 (1-1092)

: PRIOR APPLICATION NUMBER: 60/186,126
 : PRIOR FILING DATE: 2000-03-01
 : PRIOR APPLICATION NUMBER: 60/190,479
 : PRIOR FILING DATE: 2000-03-17
 : PRIOR APPLICATION NUMBER: 60/200,545
 : PRIOR FILING DATE: 2000-04-27
 : PRIOR APPLICATION NUMBER: 60/200,303
 : PRIOR FILING DATE: 2000-04-28
 : PRIOR APPLICATION NUMBER: 60/200,779
 : PRIOR FILING DATE: 2000-04-28
 : PRIOR APPLICATION NUMBER: 60/200,999
 : PRIOR FILING DATE: 2000-05-01
 : PRIOR APPLICATION NUMBER: 60/202,084
 : PRIOR FILING DATE: 2000-05-04
 : PRIOR APPLICATION NUMBER: 60/206,201

OTHER INFORMATION: n=A,T,C OR G
US-09-796-692-9477

US-09-936-737A-2 (1-103) x US-09-796-692-9477 (1-543)

RESULT 3
US-10-040-739-838
; Sequence 838, Application US/10040739
; Patent No. US20020173635A1

APPLICANT: Jacobs, Kenneth
McCoy, John
Lavalley, Edward
Racie, Lisa
Merberg, David
Treacy, Maurice
Spaulding, Vikki
TITLE OF INVENTION: SECRETED, EXPRESSED SEQUENCE TAGS
NUMBER OF SEQUENCES: 1519
CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.
 STREET: 87 Cambridgepark Drive
 CITY: Cambridge
 STATE: Massachusetts
 COUNTRY: U.S.A
 ZIP: 02140

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy Disk
 COMPUTER: IBM PC Compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/10/040,739
 FILING DATE: 07-Jan-2002
 CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 09/036,520
 FILING DATE: 03-JUN-1998

ATTORNEY/AGENT INFORMATION:
 NAME: Brown, Scott A.
 REGISTRATION NUMBER: 32,724
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 498-8224
 TELEFAX: (617) 876-5851

INFORMATION FOR SEQ ID NO: 838:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 540 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: CDNA
 SEQUENCE DESCRIPTION: SEQ ID NO: 838:
 US-10-040-739-838

Alignment Scores:
 Pred. No.: 0.866 Length: 540
 Score: 70.00 Matches: 31
 Percent Similarity: 43.75% Conservative: 18
 Best Local Similarity: 27.68% Mismatches: 50
 Query Match: 12.22% Indels: 13
 Gaps: 5

US-09-936-737a-2 (1-103) x US-10-040-739-838 (1-540)

QY 2 GUUAGGUUASPCYSTRPThrPheTyr-----AlaAsnArgLysTyrThrAsp-PH 18
 DB 204 GAAAAAACAATGTTCAACATCTACAGACTCTTCTCGAAGAAAAAACAAGACATT 263
 QY 18 eAspLysSerPheLysLys-----SerSerAspLeuAspLysLysLysThrCysPh 36
 DB 264 AGACTCAGGACACTCTGAAATTGGAATCTCCAGAAATCGAGTGTCTTAAGACA----- 318
 QY 36 eLysThrGluTyrCysTyrIleVal---PheGluAspThrValAsnLysGluCysTyrTyr 55
 DB 319 -AGAGAGAAATGCAATCAGTGTCAACTCATATATGACAAAGACAGTGTATGA 377
 QY 55 rAsnValValAspGluGluGluLeuAspGluGluLysPheValValAspGluAsnPherh 75
 DB 378 AAACCTCAACACATGATGTAATGTGTGACAAAGATGCTGTGAGGAAATTCAGAAAGCGT 437
 QY 75 rGluAsnTyrLeuThrAspCysGluGluLysAsp-----AlaGluAsnAlaAl 91
 DB 438 TAGTGAATAGCAAGTGTATGAGAAATCTGAAATGAATGAATTAATCAAGTGTGTGAGAGCTTC 497
 QY 91 aGlyThrGluLysAspLysSerAspGluValAspGlu 102
 DB 498 AGGTGATGACGATGAGATGAGATGATGATGAAGAG 531

RESULT 4
 US-09-815-242-7425 Application US/09815242
 ; Sequence 7425
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert
 APPLICANT: Ohlsen, Kari L.
 APPLICANT: Zyskind, Judith W.
 APPLICANT: Wall, Daniel
 APPLICANT: Trawick, John D.
 APPLICANT: Carr, Grant J.
 APPLICANT: Yamamoto, Robert T.
 APPLICANT: Xu, H. Howard
 TITLE OF INVENTION: Identification of Essential Genes in
 FILE REFERENCE: ELITRA 011A
 CURRENT APPLICATION NUMBER: US/09/815,242
 PRIOR APPLICATION NUMBER: 60/191,078
 PRIOR FILING DATE: 2000-03-21
 PRIOR APPLICATION NUMBER: 60/206,848
 PRIOR FILING DATE: 2000-05-23
 PRIOR APPLICATION NUMBER: 60/207,727
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: 60/242,578
 PRIOR FILING DATE: 2000-10-23
 PRIOR APPLICATION NUMBER: 60/253,625
 PRIOR FILING DATE: 2000-11-27
 PRIOR APPLICATION NUMBER: 60/257,931
 PRIOR FILING DATE: 2000-12-22
 PRIOR APPLICATION NUMBER: 60/269,308
 PRIOR FILING DATE: 2001-02-16
 NUMBER OF SEQ ID NOS: 14110
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 7425
 LENGTH: 3504
 TYPE: DNA
 ORGANISM: Helicobacter pylori
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (1)...(3504)
 US-09-815-242-7425

Alignment Scores:
 Pred. No.: 11.5 Length: 3504
 Score: 70.00 Matches: 24
 Percent Similarity: 46.15% Conservative: 12
 Best Local Similarity: 30.77% Mismatches: 24
 Query Match: 12.22% Indels: 18
 Gaps: 4

US-09-936-737a-2 (1-103) x US-09-815-242-7425 (1-3504)

QY 19 AspLysSerPheLysLysSerSerAspLeuAspGluCysLysLysThrCysPheLysThr 38
 DB 718 GATTAACACAGCAATCCCTCAACATCCCAAGAAAGCCCAAAACCTTTTGACGCT 777
 QY 39 GluTyrCysTyrIleValPheGluAspThrValAsnLysGluCysTyrTyrAsnValVal 58
 DB 778 -----ATCTTGTCGCAAGATCTAGCGAAC-----GCCGTGATTAACGTCAAG 819
 QY 59 AspGluGluGluLeuAspGluGluLysPheValValAspGluAsnPherhGluAsnTyr 78
 DB 820 CCCACGCAAA-----TTAGGGACAGCAATTAATGGAATAATTTTC 858
 QY 79 LeuThrAspCysGluGluLysAspAlaGluAsnAlaAlaGlyThrGluLysPglu 96
 DB 859 -----GCTAATAAAACGGGCAACATCGCAAGACCTTGAATGAG 897

RESULT 5
 US-09-822-246-3/c
 ; Sequence 3, Application US/09822246
 ; Patent No. US20020142383A1
 ; GENERAL INFORMATION:
 ; APPLICANT: MERKIDOV, Gennady et al.
 ; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
 ; NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
 ; AND USES THEREOF

FILE REFERENCE: CL001149
CURRENT APPLICATION NUMBER: US/09/822,246
CURRENT FILING DATE: 2001-04-02
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 3
LENGTH: 197997
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(197997)
OTHER INFORMATION: n = A,T,C or G
US-09-822-246-3

Alignment Scores:
Pred. No.: 3.04e+03 Length: 197997
Score: 70.00 Matches: 16
Percent Similarity: 50.00% Conservative: 4
Best Local Similarity: 40.00% Mismatches: 20
Query Match: 12.22% Indels: 0
DB: 10 Gaps: 0

US-09-936-737A-2 (1-103) x US-09-822-246-3 (1-197997)

QY 8 ThrPheTyAlaAsnArgLysTyThrAspPheAspLysSerPheLysLysSerSerAsp 27
||||| :||| ||| ||| :||| ||| |||
Db 182546 ACATTATGACTTTGAGCACTTTACAACTTTCACAAAGCAGGAGGAGAAAGTGAAGC 182487

QY 28 LeuSpGLucLysLysThCysPheLysThCgluTyrcysTyrlLevalPhegluasp 47
||||| :||| ||| ||| :||| ||| |||
Db 182486 CCTGATGAGTGGAAATCATTTCACTGATCATTAATCAATTTTGAAGAC 182427

RESULT 6
US-09-765-272-199
Sequence 199, Application US/09765272
Patent No. US20020061545A1

GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/765,272
FILING DATE: 22-Jan-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 199:
SEQUENCE CHARACTERISTICS:
LENGTH: 2023 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 199:
US-09-765-272-199

Alignment Scores:
Pred. No.: 6.29 Length: 2023
Score: 69.50 Matches: 25
Percent Similarity: 39.62% Conservative: 17
Best Local Similarity: 23.58% Mismatches: 35
Query Match: 12.13% Indels: 29
DB: 10 Gaps: 5

US-09-936-737A-2 (1-103) x US-09-765-272-199 (1-2023)

QY 7 TrpThrPheTyAlaAsnArgLysTyThrAspPheAspLysSerPheLysLysSerSer 26
||||| :||| ||| ||| :||| ||| |||
Db 965 TGGATTAGCAATTAATTAAGCGCTTACAGAGATTGAC-----TCA 1006

QY 27 AspLeuAspGLucLysLysThr-----CysPheLysThCgluTyrcys 41
:||||| :||| ||| :||||| :||| |||
Db 1007 GAATATATTAAATTGTCGGGTACTCTGAGCAATTTTCATGTTAAATCGAATATAGT 1066

QY 42 TyrlLevalPhegluaspThrValasnLysglucysTyTyrrasnValvalaspLysglu 61
||||| :||| ||| ||| :||| ||| |||
Db 1067 TACACAGTCTTT-----TTACGCTATTTCATGAGTGAATTCGTG 1105

QY 62 GluLeuAspGLuLysPhe-----ValvalaspLysPheThrglu 76
:||||| :||| ||| :||||| :||| |||
Db 1106 CAAAGAGACAGAGCCCTCTACTTGACCTGTGATCTAGTTGTTACGAAATACTGATGAC 1165

QY 77 AsnTyrlLeuThrAspCysgluLys-----LysaspLysAla 90
:||||| :||| ||| :||||| :||| |||
Db 1166 TTGTTGCTACAGACTTACAGATTATCCTTTGCTGCTGTAGAGATTTGGGGCAGA 1225

QY 91 AlaGlyThrgLysaspLys 96
||||| :||| ||| ||| :||| ||| |||
Db 1226 GCTTATTTTGGTCAAGAA 1243

RESULT 7
US-09-729-674-129/C
Sequence 129, Application US/09729674
Patent No. US20010039335A1

GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
APPLICANT: McCoy, John M.
APPLICANT: Lavallee, Edward R.
APPLICANT: Collins-Racie, Lisa A.
APPLICANT: Evans, Cheryl
APPLICANT: Merberg, David
APPLICANT: Treacy, Maurice
APPLICANT: Agostino, Michael J.
APPLICANT: Steindinger II, Robert J.
APPLICANT: Spaulding, Vikki
APPLICANT: Wong, Gordon G.
APPLICANT: Clark, Hilary
APPLICANT: Fechtel, Kim

TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
FILE REFERENCE: 6055-64X
CURRENT APPLICATION NUMBER: US/09/729,674
FILING DATE: 2000-12-04
PRIOR APPLICATION NUMBER: 09/539,330
PRIOR FILING DATE: 2000-03-30
NUMBER OF SEQ ID NOS: 283
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO: 129
LENGTH: 1861
TYPE: DNA
ORGANISM: Homo sapiens
US-09-729-674-129

Alignment Scores:
Pred. No.: 8.98 Length: 1861
Score: 68.00 Matches: 27

Percent Similarity: 40.00%
 Best Local Similarity: 25.71%
 Query Match: 11.87%
 DB: 10
 Conservative: 15
 Mismatches: 41
 Indels: 22
 Gaps: 5

US-09-936-737a-2 (1-103) x US-09-729-674-129 (1-1861)

QY 9 PheTyrAlaAsnArgLysTyr-----ThraspPheasp 19
 |||:|||||
 DB 1656 TTTGGCAATCGGAATATCATATGAGCCCATTCGCCAACAAGAGACTGAAATCAT 1597
 |||:|||||

QY 20 LysSerPheLysSerSerAspLeuAspLysLysLysThrCysPheLys-----37
 |||:|||||
 DB 1396 ACGAGCAAGCCAAAGCTTCAGATTGGA---GCCAGATGACCATGATCAATGACAGA 1540
 |||:|||||

QY 38 -----ThrgLutyrCys-----TyrileValPheGluAspThr 48
 |||:|||||
 DB 1539 AGATYAAAGAAAGATGATCATGAGAGTTCTGTACACATTATTAATGCTCTTCAGATTCA 1480
 |||:|||||

QY 49 ValAsnLysGluCysTyrTyrAsnValValAspLysGluLysAspGlnGluLysPhe 68
 |||:|||||
 DB 1479 ATTACATTTCTTGC---ACCACTGTGTCAAGTTAAATCAAGAAGCAACGCGATT 1423
 |||:|||||

QY 69 ValValAspGluAsnPheThrGluAsnTyrLeuThrAspCysGluLysAspAlaGly 88
 |||:|||||
 DB 1422 CTACACCAATGACCGCTGCACAAACGACGACGATGATGCGACAGATTGCGA 1363
 |||:|||||

QY 89 AsnAlaAlaGlyThr 93
 |||:|||||

DB 1362 CTGTCTCCGGGACA 1348

RESULT 8
 US-09-864-761-4740
 ; Sequence 4740, Application US/09864761
 ; Patent No. US20020048763A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Penn, Sharon G.
 ; APPLICANT: Rank, David R.
 ; APPLICANT: Hanzel, David K.
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 ; FILE REFERENCE: Aemlica-x-1
 ; CURRENT APPLICATION NUMBER: US/09/864,761
 ; CURRENT FILING DATE: 2001-05-23
 ; PRIOR APPLICATION NUMBER: US 60/180,312
 ; PRIOR FILING DATE: 2000-02-04
 ; PRIOR APPLICATION NUMBER: US 60/207,456
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: US 09/632,366
 ; PRIOR FILING DATE: 2000-08-03
 ; PRIOR APPLICATION NUMBER: GB 24263,6
 ; PRIOR FILING DATE: 2000-10-04
 ; PRIOR APPLICATION NUMBER: US 60/236,359
 ; PRIOR FILING DATE: 2000-09-27
 ; PRIOR APPLICATION NUMBER: PCT/US01/00666
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00667
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00664
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00669
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00665
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00668
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00663
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00662
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00661
 ; PRIOR FILING DATE: 2001-01-30

;; PRIOR APPLICATION NUMBER: PCT/US01/00670
 ;; PRIOR FILING DATE: 2001-01-30
 ;; PRIOR APPLICATION NUMBER: US 60/234,687
 ;; PRIOR FILING DATE: 2000-09-21
 ;; PRIOR APPLICATION NUMBER: US 09/608,408
 ;; PRIOR FILING DATE: 2000-06-30
 ;; PRIOR APPLICATION NUMBER: US 09/774,203
 ;; PRIOR FILING DATE: 2001-01-29
 ;; NUMBER OF SEQ ID NOS: 49117
 ;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
 ;; SEQ ID NO 4740
 ;; LENGTH: 1988
 ;; TYPE: DNA
 ;; ORGANISM: Homo sapiens
 ;; FEATURE:
 ;; OTHER INFORMATION: MAP TO AC002086.1
 ;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1
 ;; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.5
 ;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.5
 ;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.4
 ;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3
 ;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.8
 ;; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.1
 ;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1
 ;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2
 ;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.6
 US-09-864-761-4740

Alignment Scores:
 Pred. No.: 9.84 Length: 1988
 Score: 68.00 Matches: 24
 Percent Similarity: 46.99% Conservative: 15
 Best Local Similarity: 28.92% Mismatches: 34
 Query Match: 11.87% Indels: 10
 DB: 10 Gaps: 2

US-09-936-737a-2 (1-103) x US-09-864-761-4740 (1-1988)

QY 1 GluGluArgGluAspCysTyr-ThraspPheaspLysThrAspPheaspLys 20
 |||:|||||
 DB 905 AAGGAAGTGAATAATGTTGATCTTCACTCCACACA-----TGTGACTCCCAATA 958
 |||:|||||

QY 20 sSerPheLysSerSerAspLeuAspLysLysLysThrCysPheLysThrGluTyr 40
 |||:|||||
 DB 959 TATTTGTTAATACAGACACACTTCTACACCAACCAAAAGTCACTTCACTTCCA 1018
 |||:|||||

QY 40 rCysTyrileValPhe-----GluAspThrValAsnLysGluCysTyr 54
 |||:|||||
 DB 1019 ATCATATGCCCTCTCAATCAATTAATCTTGTGACGAATACACAGACCAACAGTGCTA 1078
 |||:|||||

QY 54 rTyrAsnValValAspGlyLys-GluLeuAspGlnGluLysPheValValAspGluAsp 74
 |||:|||||
 DB 1079 TTTTACAGGAACAAGCCCAATGTAAGAGGAGGAGGAATATATGATGATGATG 1138
 |||:|||||

QY 74 heThr 75
 |||

DB 1139 ACACCT 1143

RESULT 9
 US-09-727-384-3
 ; Sequence 3, Application US/09727384
 ; Patent No. US20020098511A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Myriad Genetics, Inc.
 ; APPLICANT: Heichman, Karen
 ; APPLICANT: Cimbora, Daniel M.
 ; APPLICANT: Bush, Angie
 ; APPLICANT: Mauck, Kimberly
 ; APPLICANT: Bartel, Paul L.
 ; TITLE OF INVENTION: Protein-Protein Interactions
 ; FILE REFERENCE: 2318-271
 ; CURRENT APPLICATION NUMBER: US/09/727,384
 ; CURRENT FILING DATE: 2000-12-01

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

```

;
;      TYPE: nucleic acid
;      STRANDEDNESS: single
;
;      TOPOLOGY: linear
;
;      MOLECULE TYPE: Genomic DNA
;      SEQUENCE DESCRIPTION: SEQ ID NO: 257:
US-09-939-980-257

```

Alignment scores:	
Pred. No.:	25
Length:	3110

Score:	07.00	Matches:	22
Percent Similarity:	42.50%	Conservative:	19

Best Local Similarity:	20.0%	Indels:	30
Query Match:	11.69%	Indels:	7

DB: 100-330-090-257 (1-3110)

Metabolism of the amino acid sequence Ser¹His²Arg³Leu⁴Val⁵Met⁶Leu⁷Val⁸Met⁹Leu¹⁰Val¹¹Met¹²Leu¹³Val¹⁴Met¹⁵Leu¹⁶Val¹⁷Met¹⁸Leu¹⁹Val²⁰Met²¹Leu²²Val²³Met²⁴Leu²⁵Val²⁶Met²⁷Leu²⁸Val²⁹Met³⁰Leu³¹Val³²Met³³Leu³⁴Val³⁵Met³⁶Leu³⁷Val³⁸Met³⁹Leu⁴⁰Val⁴¹Met⁴²Leu⁴³Val⁴⁴Met⁴⁵Leu⁴⁶Val⁴⁷Met⁴⁸Leu⁴⁹Val⁵⁰Met⁵¹Leu⁵²Val⁵³Met⁵⁴Leu⁵⁵Val⁵⁶Met⁵⁷Leu⁵⁸Val⁵⁹Met⁶⁰Leu⁶¹Val⁶²Met⁶³Leu⁶⁴Val⁶⁵Met⁶⁶Leu⁶⁷Val⁶⁸Met⁶⁹Leu⁷⁰Val⁷¹Met⁷²Leu⁷³Val⁷⁴Met⁷⁵Leu⁷⁶Val⁷⁷Met⁷⁸Leu⁷⁹Val⁸⁰Met⁸¹Leu⁸²Val⁸³Met⁸⁴Leu⁸⁵Val⁸⁶Met⁸⁷Leu⁸⁸Val⁸⁹Met⁹⁰Leu⁹¹Val⁹²Met⁹³Leu⁹⁴Val⁹⁵Met⁹⁶Leu⁹⁷Val⁹⁸Met⁹⁹Leu¹⁰⁰Val¹⁰¹Met¹⁰²Leu¹⁰³Val¹⁰⁴Met¹⁰⁵Leu¹⁰⁶Val¹⁰⁷Met¹⁰⁸Leu¹⁰⁹Val¹¹⁰Met¹¹¹Leu¹¹²Val¹¹³Met¹¹⁴Leu¹¹⁵Val¹¹⁶Met¹¹⁷Leu¹¹⁸Val¹¹⁹Met¹²⁰Leu¹²¹Val¹²²Met¹²³Leu¹²⁴Val¹²⁵Met¹²⁶Leu¹²⁷Val¹²⁸Met¹²⁹Leu¹³⁰Val¹³¹Met¹³²Leu¹³³Val¹³⁴Met¹³⁵Leu¹³⁶Val¹³⁷Met¹³⁸Leu¹³⁹Val¹⁴⁰Met¹⁴¹Leu¹⁴²Val¹⁴³Met¹⁴⁴Leu¹⁴⁵Val¹⁴⁶Met¹⁴⁷Leu¹⁴⁸Val¹⁴⁹Met¹⁵⁰Leu¹⁵¹Val¹⁵²Met¹⁵³Leu¹⁵⁴Val¹⁵⁵Met¹⁵⁶Leu¹⁵⁷Val¹⁵⁸Met¹⁵⁹Leu¹⁶⁰Val¹⁶¹Met¹⁶²Leu¹⁶³Val¹⁶⁴Met¹⁶⁵Leu¹⁶⁶Val¹⁶⁷Met¹⁶⁸Leu¹⁶⁹Val¹⁷⁰Met¹⁷¹Leu¹⁷²Val¹⁷³Met¹⁷⁴Leu¹⁷⁵Val¹⁷⁶Met¹⁷⁷Leu¹⁷⁸Val¹⁷⁹Met¹⁸⁰Leu¹⁸¹Val¹⁸²Met¹⁸³Leu¹⁸⁴Val¹⁸⁵Met¹⁸⁶Leu¹⁸⁷Val¹⁸⁸Met¹⁸⁹Leu¹⁹⁰Val¹⁹¹Met¹⁹²Leu¹⁹³Val¹⁹⁴Met¹⁹⁵Leu¹⁹⁶Val¹⁹⁷Met¹⁹⁸Leu¹⁹⁹Val²⁰⁰Met²⁰¹Leu²⁰²Val²⁰³Met²⁰⁴Leu²⁰⁵Val²⁰⁶Met²⁰⁷Leu²⁰⁸Val²⁰⁹Met²¹⁰Leu²¹¹Val²¹²Met²¹³Leu²¹⁴Val²¹⁵Met²¹⁶Leu²¹⁷Val²¹⁸Met²¹⁹Leu²²⁰Val²²¹Met²²²Leu²²³Val²²⁴Met²²⁵Leu²²⁶Val²²⁷Met²²⁸Leu²²⁹Val²³⁰Met²³¹Leu²³²Val²³³Met²³⁴Leu²³⁵Val²³⁶Met²³⁷Leu²³⁸Val²³⁹Met²⁴⁰Leu²⁴¹Val²⁴²Met²⁴³Leu²⁴⁴Val²⁴⁵Met²⁴⁶Leu²⁴⁷Val²⁴⁸Met²⁴⁹Leu²⁵⁰Val²⁵¹Met²⁵²Leu²⁵³Val²⁵⁴Met²⁵⁵Leu²⁵⁶Val²⁵⁷Met²⁵⁸Leu²⁵⁹Val²⁶⁰Met²⁶¹Leu²⁶²Val²⁶³Met²⁶⁴Leu²⁶⁵Val²⁶⁶Met²⁶⁷Leu²⁶⁸Val²⁶⁹Met²⁷⁰Leu²⁷¹Val²⁷²Met²⁷³Leu²⁷⁴Val²⁷⁵Met²⁷⁶Leu²⁷⁷Val²⁷⁸Met²⁷⁹Leu²⁸⁰Val²⁸¹Met²⁸²Leu²⁸³Val²⁸⁴Met²⁸⁵Leu²⁸⁶Val²⁸⁷Met²⁸⁸Leu²⁸⁹Val²⁹⁰Met²⁹¹Leu²⁹²Val²⁹³Met²⁹⁴Leu²⁹⁵Val²⁹⁶Met²⁹⁷Leu²⁹⁸Val²⁹⁹Met³⁰⁰Leu³⁰¹Val³⁰²Met³⁰³Leu³⁰⁴Val³⁰⁵Met³⁰⁶Leu³⁰⁷Val³⁰⁸Met³⁰⁹Leu³¹⁰Val³¹¹Met³¹²Leu³¹³Val³¹⁴Met³¹⁵Leu³¹⁶Val³¹⁷Met³¹⁸Leu³¹⁹Val³²⁰Met³²¹Leu³²²Val³²³Met³²⁴Leu³²⁵Val³²⁶Met³²⁷Leu³²⁸Val³²⁹Met³³⁰Leu³³¹Val³³²Met³³³Leu³³⁴Val³³⁵Met³³⁶Leu³³⁷Val³³⁸Met³³⁹Leu³⁴⁰Val³⁴¹Met³⁴²Leu³⁴³Val³⁴⁴Met³⁴⁵Leu³⁴⁶Val³⁴⁷Met³⁴⁸Leu³⁴⁹Val³⁵⁰Met³⁵¹Leu³⁵²Val³⁵³Met³⁵⁴Leu³⁵⁵Val³⁵⁶Met³⁵⁷Leu³⁵⁸Val³⁵⁹Met³⁶⁰Leu³⁶¹Val³⁶²Met³⁶³Leu³⁶⁴Val³⁶⁵Met³⁶⁶Leu³⁶⁷Val³⁶⁸Met³⁶⁹Leu³⁷⁰Val³⁷¹Met³⁷²Leu³⁷³Val³⁷⁴Met³⁷⁵Leu³⁷⁶Val³⁷⁷Met³⁷⁸Leu³⁷⁹Val³⁸⁰Met³⁸¹Leu³⁸²Val³⁸³Met³⁸⁴Leu³⁸⁵Val³⁸⁶Met³⁸⁷Leu³⁸⁸Val³⁸⁹Met³⁹⁰Leu³⁹¹Val³⁹²Met³⁹³Leu³⁹⁴Val³⁹⁵Met³⁹⁶Leu³⁹⁷Val³⁹⁸Met³⁹⁹Leu⁴⁰⁰Val⁴⁰¹Met⁴⁰²Leu⁴⁰³Val⁴⁰⁴Met⁴⁰⁵Leu⁴⁰⁶

1182 ACCGACCTTTTACCAATAGACCAACGATGTGTATTTT-----CGATTTCAGGCAAAACATCAGG 1236

38 Leu---AspG1 uCvslvslvsthrCvSpheIvsthrGluTfrcYstYrrIleValpheGlu 46

1237 TTCTGGAACGATGCCACAACGTTACTATTAATAATTATGGGTCGAATATGAAA 1296

47 AsptHrValAsnLvsGlu-----CysTyrTrpAsnVal----- 57

Db 1297 ATACAGTAACGATAAAACGAAGTTATCGGATACGTTAATACTGGCGTTTACCAAT 1356

QY	58	-----ValaspGly-----	-----GlulGluleuaspGlnGlu	66
----	----	---------------------	--------------------------	----

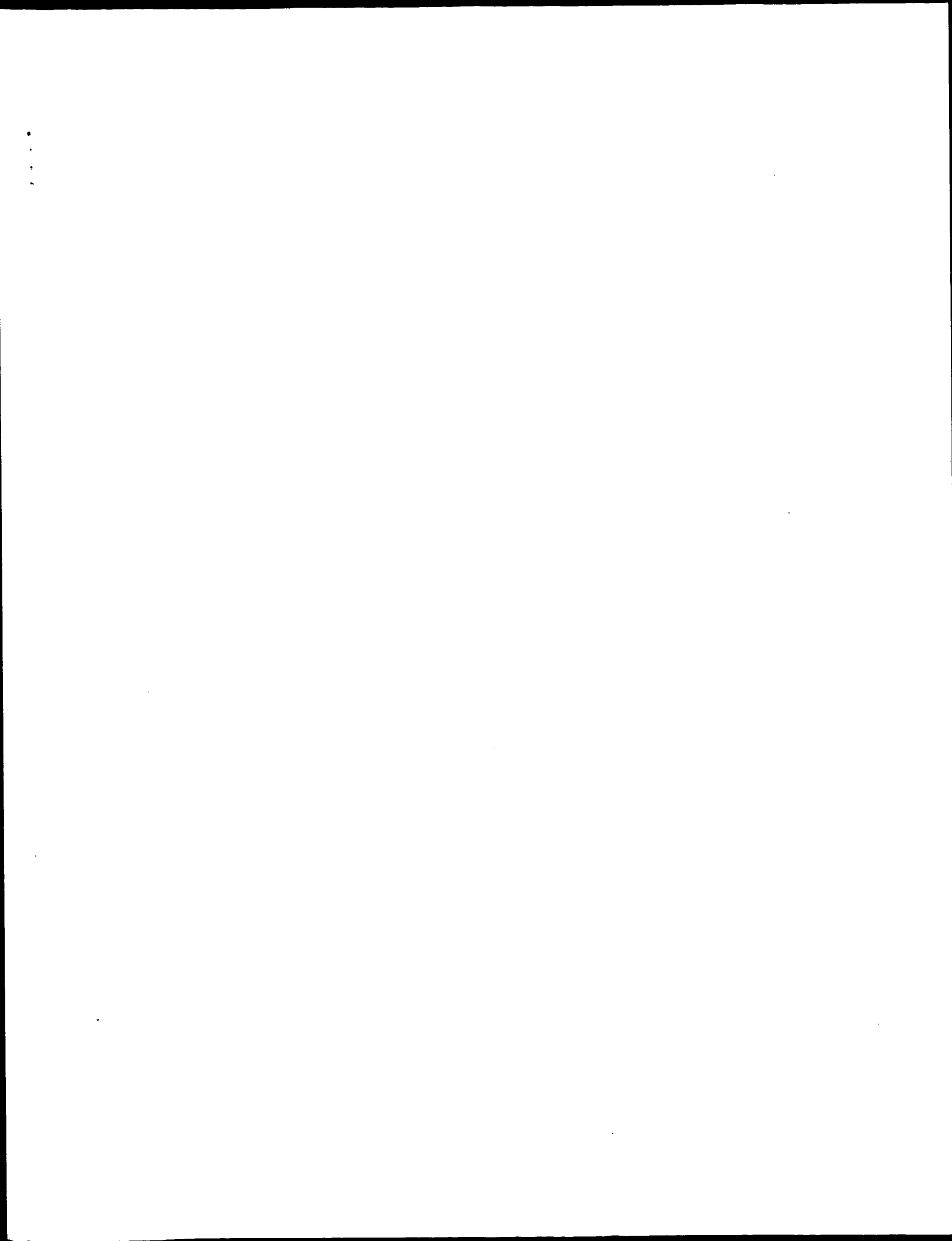
Wed Feb 26 15:59:05 2003

us-09-936-737a-2.rnpb

Page 9

Db 1357 AGTTTAATGATGATGATATACCAATGTGGCTATCAATTCGAAGAAGACTTTGAACCTAGA 1416
 QY 67 LysDeValValaspGluSnpherhrglu-----AsnTyrLeuThrAspGlu 83
 Db 1417 AAGTTGTT-----TTCACTAACGGCGGAATTAATATATAACGCAATTTTGA 1464
 QY 84 GlyLysAspAlaGlnAlaAlaGlyThrGlyAspGluSerAspGluValAspGluAsp 103
 Db 1465 AAGAGAACGCTACCGCATGCATCAACGCCAACAAGATGATCGATTGAGATGATGAGAA 1524

```
Search completed: February 25, 2003, 03:57:32
Job time : 88 secs
```



GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 25, 2003, 02:29:55 ; Search time 14. Seconds

(without alignments)
707.275 Million cell updates/sec

Title: US-09-936-737a-2

Perfect score: 573
Sequence: 1 EEREDCWTFFANRKYTFDFK.....GKDAGNAGTGDESDVEDD 103

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues.

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR_73:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query length	DB ID	Description
1	80.5	14.0	147	2	A42435 leech antiplatelet
2	80.5	14.0	412	2	A55320 immunophilin FKBP4
3	78.5	13.7	550	2	T26562 hypothetical prote
4	78.5	13.7	907	2	T27317 hypothetical prote
5	78	13.6	682	2	B84021 two-component sens
6	78	13.6	797	2	A36811 hypothetical prote
7	75.5	13.2	264	2	T22054 hypothetical prote
8	75.5	13.2	491	2	T16354 hypothetical prote
9	75	13.1	360	2	T34510 hypothetical prote
10	75	13.1	3844	2	T18402 asparagine/asparat
11	74.5	13.0	1277	2	T32731 PAR interacting pr
12	74	12.9	296	2	D70465 hypothetical prote
13	74	12.9	1278	2	B70236 hypothetical prote
14	74	12.9	4466	1	S17231 dynein beta heavy
15	73	12.7	643	2	T21428 hypothetical prote
16	73	12.7	1405	1	D72PA DNA-directed DNA p
17	73	12.7	2543	2	T31687 surface antigen - p
18	72.5	12.7	564	2	T25945 hypothetical prote
19	72.5	12.7	2533	2	T28675 alpha-5ID immobill
20	72.5	12.7	2533	2	T28675 alpha-5ID immobill
21	72	12.6	1503	2	B84493 probable Athila re
22	71.5	12.5	320	2	A71611 Ser/Thr protein ki
23	71.5	12.5	594	2	F81359 para-aminobenzoate
24	71	12.4	782	2	S50719 hypothetical prote
25	70.5	12.3	146	1	T37222 phospholipase A2 (
26	70.5	12.3	156	2	S60109 nuclear cap bindin
27	70.5	12.3	156	2	G96774 hypothetical prote
28	70.5	12.3	256	2	G96774 hypothetical prote
29	70.5	12.3	748	2	C64409 hypothetical prote

30	70.5	12.3	786	2	C86406 88.6K hypothetical
31	70	12.2	328	2	T24432 hypothetical prote
32	70	12.2	1058	2	S50295 hypothetical prote
33	70	12.2	1167	2	F71909 hypothetical prote
34	70	12.2	1255	2	B97104 hypothetical prote
35	69.5	12.1	146	1	PSR6A phospholipase A2 (
36	69.5	12.1	452	2	T37899 hypothetical prote
37	69.5	12.1	628	2	S61160 hypothetical prote
38	69.5	12.1	696	2	D95206 hypothetical prote
39	69.5	12.1	738	2	A13625 ribonucleoside-dip
40	69	12.0	164	2	T32824 hypothetical prote
41	69	12.0	374	2	T33328 hypothetical prote
42	69	12.0	1127	2	T28317 ORF MSY156 hypothe
43	69	12.0	1650	2	T18444 hypothetical prote
44	68.5	12.0	385	2	T26487 hypothetical prote
45	68.5	12.0	597	2	D70100 phosphoglucomutase

ALIGNMENTS

```

RESULT 1
A42435 leech antiplatelet protein precursor - Mexican leech
C:Species: Hementeria officinalis (Mexican leech)
C>Date: 07-Apr-1994 #Sequence_Revision 07-Apr-1994 #text_change 07-May-1999
C:Accession: A42435; A42434
R:Keller, P.M.; Schultz, L.D.; Condra, C.; Karczewski, J.; Connolly, T.M.
J. Biol. Chem. 267, 6899-6904, 1992
A:Title: An inhibitor of collagen-stimulated platelet activation from the salivary gl
A:Reference number: A42435; MUID:92202247; PMID:1551898
A:Accession: A42435
A:Molecule type: mRNA
A:Residues: 1-147 <KE>
A:Cross-References: GB:M81489
A:Note: the authors translated the codon AGC for residue 20 as Arg and GAA for residu
A:Note: the authors' translation is shown at residue 65
A:Note: sequence extracted from NCBI backbone (NCBI:91615) and modified
R:Connolly, T.M.; Jacobs, J.W.; Condra, C.
J. Biol. Chem. 267, 6893-6898, 1992
A:Title: An inhibitor of collagen-stimulated platelet activation from the salivary gl
A:Reference number: A42434; MUID:92202246; PMID:1551897
A:Accession: A42434
A:Molecule type: protein
A:Residues: 59-91;101-113;123-139 <CON>
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-147/Product: leech antiplatelet protein #status predicted <MAT>

Query Match 14.0%; Score 80.5; DB 2; Length 147;
Best Local Similarity 32.7%; Pred. No. 1.7;
Matches 18; Conservative 10; Mismatches 24; Indels 3; Gaps 1;

QY 5 DCWTFYANRKYTFDFKSFKSSDLECKKCTFETRY---CYIVEDYVKNCCYYN 56
Db 66 DCMWRKRPWKLPDNLTKTEFTSVDECKRMCESAVEPSCYILQINTEFTNCCYYN 120
A55320
RESULT 2
A55320 immunophilin FKBP46 - fall armyworm
C:Species: Spodoptera frugiperda (fall armyworm)
C>Date: 06-Feb-1995 #Sequence_Revision 06-Feb-1995 #text_change 02-Sep-2000
C:Accession: A55320
R:Allemri, E.S.; Fernandes-Alnemri, T.; Pometenke, K.; Robertson, N.M.; Dudley, K.; D
J. Biol. Chem. 269, 30828-30834, 1994
A:Title: FKBP46, a novel S19 insect cell nuclear immunophilin that forms a protein-ki
A:Reference number: A55320; MUID:95074110; PMID:7527037
A:Accession: A55320
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-412 <ALN>
A:Cross-References: GB:U15038; NID:g9595844; PID:g9595845
C:Superfamily: yeast peptidylprolyl isomerase FPR3; BKBP-type peptidylprolyl isomerases

```

C:Keywords: nucleus; phosphoprotein
E:324-371/Domains: BKB-type peptidylprolyl isomerase homology <PPI>

Query Match 14.0%; Score 80.5; DB 2; Length 412;
Best Local Similarity 30.6%; Pred. No. 4.7;

Matches 30; Conservative 12; Mismatches 29; Indels 27; Gaps 5;

QY 11 ANRKYTFDPKSKFKS-----SDLECKKTCFTECYIVFEDTVKKECYNNVDSGEELDQ 65

DB 133 ANKKAAPDKKAGKNSAPAEASDSDDD-----EDQLK-----FLDGEDIDTF 174

QY 66 EKPVNDEFNTYLTDFCEKGDAGNAAGTDESDEVED 103

DB 175 DE--NDESEKMN--TSABEDSDSEDDDEDEDEDD 208

RESULT 3
T26562
hypothetical protein Y26DA.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T26562
R:White, S.

submitted to the EMBL Data Library, September 1999

A:Reference number: Z20234

A:Accession: T26562

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-550 <WIL>
A:Cross-references: EMBL:AL110478; NID:el542139; PIDN:CAB54340.1; CESP:Y26DA.2

A:Experimental source: clone Y26DA4

A:Gene: CESP:Y26DA.2

A:Introns: 134/3; 160/1; 286/3; 316/1; 480/2; 502/2

Query Match 13.7%; Score 78.5; DB 2; Length 550;
Best Local Similarity 25.6%; Pred. No. 9.7;

Matches 23; Conservative 12; Mismatches 32; Indels 23; Gaps 3;

QY 6 CWTFFYANRKYTFD-----DKSEKKS-SDLECKKTCFTECYIVF 43

DB 170 CGNLFHRRPYRPSCECTECMDTQVNAKPGPLMGSRNKNILNLSDECVKCYCKMDLNCV 229

QY 44 VFEDTVKKECYNNVDSGEELDQEKFYVDEN 73

DB 230 NYDKSKCKEWMWSIDVNH-LEKVFHSEN 258

RESULT 4
T27317
hypothetical protein Y69H2.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T27317
R:McMurray, A.

submitted to the EMBL Data Library, August 1997

A:Reference number: Z20343

A:Accession: T27317

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-907 <WIL>
A:Cross-references: EMBL:Z98877; PIDN:CAB11570.1; GSPDB:GN00023; CESP:Y69H2.2

A:Experimental source: clone Y69H2
A:Gene: CESP:Y69H2.2

A:Map position: 5
A:Introns: 23/3; 65/1; 105/1; 146/1; 191/1; 241/1; 293/1; 381/1; 419/1; 461/1; 499/1; 55

Query Match 13.7%; Score 78.5; DB 2; Length 907;
Best Local Similarity 29.2%; Pred. No. 16;

Matches 33; Conservative 12; Mismatches 33; Indels 35; Gaps 8;

QY 10 YANRKYTFDPKSKFKS-----KSSDL-----DECKKTCFTECYIVFEDT 48

DB 226 YAQKSTP-DKTLKYTLFCGECWCVSSEMDLVKDDVSKMKDCKKSCGGDSSEYC-----GK 279

QY 49 VNRKCYNNVDSGEELDQEKFYVDENFTYLTDFCEKGDAGNAAGTGDSEVED 101

DB 280 VNSRCL--VEEAGEME-----TDPNACEDNSTLC-GAELSHGCMCKINQSDVTD 324

RESULT 5

B84021
two-component sensor histidine kinase involved in chemotaxis cheA [imported] - Bac111

C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001

C:Accession: B84021

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; H

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a

A:Reference number: A83650; MUID:20512582; PMID:11058132

A:Accession: B84021

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-682 <SNO>

A:Cross-references: GB:AP001517; GB:BA000004; NID:910175500; PIDN:BA006689.1; GSPDB:G

A:Experimental source: strain C-125

A:Gene: cheA

C:Superfamily: chemotaxis protein cheA

Query Match 13.6%; Score 78; DB 2; Length 682;
Best Local Similarity 32.6%; Pred. No. 13;

Matches 28; Conservative 13; Mismatches 31; Indels 14; Gaps 5;

QY 10 YANRKYTFDPKSKFKS-----LDECKKTCFTECYIVF--DTVNKCYNNVDS 60

DB 156 YDFEFTVLDQSEFGYQAYQIEVTLDE--KTLKAAKRVFVFLBOVG-EVIKSTPSA 212

QY 61 EELDQEKFYVDENFTYLTDFCEKGD 86

DB 213 EELDEEKF--DERFLVTLTKVDGEE 236

RESULT 6
A36811
hypothetical protein ORF48 - saimiriine herpesvirus 1 (strain 11)

C:Species: saimiriine herpesvirus 1
A:Note: host Saimiri sciureus (common squirrel monkey)

C:Date: 16-Oct-1992 #sequence_revision 16-Oct-1992 #text_change 08-Oct-1999

C:Accession: A36811

R:Albrecht, J.

submitted to the EMBL Data Library, January 1992

A:Description: Primary structure of the Herpesvirus saimiri genome.

A:Reference number: A36806

A:Accession: A36811

A:Molecule type: DNA

A:Residues: 1-797 <ALB>
A:Cross-references: GB:X64346; NID:960320; PIDN:CAA45671.1; PID:960369

R:Albrecht, J.C.; Nicholas, J.; Biller, D.; Cameron, K.R.; Blesinger, B.; Newman, C.;

A:Title: Primary structure of the herpesvirus saimiri genome.
A:Reference number: A37309; MUID:92333688; PMID:1321287

A:Contents: annotation; protein-coding frames
A:Note: neither protein nor nucleotide sequence is given

C:Gene: 48

Query Match 13.6%; Score 78; DB 2; Length 797;
Best Local Similarity 32.3%; Pred. No. 16;

Matches 31; Conservative 12; Mismatches 33; Indels 20; Gaps 5;

QY 11 ANRKYTFDPKSKFKS-----LDECKKTCFTECYIVFEDTVKKECYNN-----VDSGEELDQ 66

DB 382 ANEK--EYKKIIDSDDRDCKDEYELE-----NEE--YNRDEEDEDGEDEDE 427

OY 67 KEVVDENFTENYLTDCGKAGNAAGTGDESDVD 102
 Db 428 KDEKEGEDEGDGEDEGEDEGE--DEGEDEGEDE 461

RESULT 7

T22054
 hypothetical protein F40G12.5 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Sep-2000
 C:Accession: T22054

R:McMurray, A.
 submitted to the EMBL Data Library, July 1996

A:Reference number: Z19507
 A:Accession: T22054
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-264 <TAL>
 A:Cross-references: EMBL:Z77661; PIDN:CA801183.1; GSPDB:GN00023; CESP:F40G12.5
 A:Experimental source: clone F40G12
 C:Genetics:
 A:Gene: CESP:F40G12.5
 A:Map position: 5
 A:Introns: 57/3: 158/3
 C:Superfamily: Caenorhabditis elegans hypothetical protein F40G12.4

Query Match 13.2%; Score 75.5; DB 2; Length 264;
 Best Local Similarity 27.7%; Pred. No. 8.8;
 Matches 31; Conservative 19; Mismatches 33; Indels 29; Gaps 8;

OY 1 EERECCWTFYANKRTYTDGSKK--SSPLDECKTCFTECYIVFE-----DYVKE 52
 Db 29 ESEDEC-----FEKVFLLIISGKHECSK---DYDFLARNLIQRRALTSKE 72

OY 53 CYYNVVDGEELDOEKF-VVDENFTE--NYLFDGEGKDGNAAGTGDESDVD 101
 Db 73 CFLEIVK-ECEPEKEFKLIENYSQVLTLLTE-KPKDNGACTAPYFLEETE 122

RESULT 8

T16354
 hypothetical protein F4ZG9.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans

C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
 C:Accession: T16354

R:Taich, A.
 submitted to the EMBL Data Library, March 1996

A:Description: The sequence of C. elegans cosmid F4ZG9.
 A:Reference number: Z18498
 A:Accession: T16354
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-491 <TAI>
 A:Cross-references: EMBL:U00051; NID:q1216305; PID:q1216310; PIDN:AAA91358.1; CESP:F4ZG9
 A:Experimental source: strain Bristol N2
 C:Genetics:
 A:Gene: CESP:F4ZG9.1
 A:Introns: 37/3; 254/2; 364/2; 406/2

Query Match 13.2%; Score 75.5; DB 2; Length 491;
 Best Local Similarity 24.5%; Pred. No. 16;
 Matches 23; Conservative 18; Mismatches 28; Indels 25; Gaps 4;

OY 10 YANKRYDF--DKSEKSSDLDECKTCFTECYIVFEEDTVNKECYYNVVDGEELDOEK 67
 Db 67 FTSAKLPDFLKERKFEADVAVAECLQKAF-----VDFDFIRAF-----ESKKEIK 112

OY 68 FVVDENFTENYLTDCGKAGNAAGTGDESDVD 101
 Db 113 DIGDEG-----KPKKAGGEADSEDEADRID 137

RESULT 9

T34510
 hypothetical protein ZK1290.10 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jun-2000
 C:Accession: T34510

R:Taich, A.
 submitted to the EMBL Data Library, July 1995

A:Description: The sequence of C. elegans cosmid ZK1290.
 A:Reference number: Z21535
 A:Accession: T34510
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-360 <TAI>
 A:Cross-references: EMBL:U21308; PIDN:AA893316.1; GSPDB:GN00020; CESP:ZK1290.10
 A:Experimental source: strain Bristol N2; clone ZK1290
 C:Genetics:
 A:Gene: CESP:ZK1290.10
 A:Map position: 2
 A:Introns: 23/1; 48/1; 158/3; 216/1; 264/1; 312/1; 331/3
 C:Superfamily: Caenorhabditis elegans hypothetical protein ZK1290.10

Query Match 13.1%; Score 75; DB 2; Length 360;
 Best Local Similarity 33.3%; Pred. No. 13;
 Matches 23; Conservative 9; Mismatches 23; Indels 14; Gaps 3;

OY 46 EDYVNRKCYYNVVDGEELDOEKRYV-----DENFTENYLTDCGKAGN----AAGTG 94
 Db 224 EDYVGE--NVDFDFKLPDKKIYIRNDGSTNTTEQSTTEKTKGKTIYVSVELGSG 280

OY 95 DESDEVED 103
 Db 281 DDEDEND 289

RESULT 10

T18402
 asparagine/aspartate rich protein - malaria parasite (Plasmodium falciparum) (fragmen
 C:Species: Plasmodium falciparum
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
 C:Accession: T18402

R:Barale, J.C.; Candelle, D.; Attal-Bonnefoy, G.; Dehoux, P.; Bonnefoy, S.; Ridley, R.
 Infect. Immun. 65, 3003-3010, 1997
 A:Title: Plasmodium falciparum AARP1, a giant protein containing repeated motifs rich
 A:Reference number: Z18929; MUID:97378065; PMID:9234746
 A:Accession: T18402

A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-3844 <BAR>
 A:Cross-references: EMBL:Y08926; NID:e1154302; PID:e311435; PIDN:CAA70130.1
 C:Genetics:
 A:Gene: aarp1

Query Match 13.1%; Score 75; DB 2; Length 3844;
 Best Local Similarity 23.9%; Pred. No. 1.5e+02;
 Matches 21; Conservative 23; Mismatches 34; Indels 10; Gaps 3;

OY 1 EERECCWTFYANKRTYD-----FDK-SFKSSDLDECKTCFTECYIVFEEDTVNKE 51
 Db 1726 EDEECQIKESYKKSECNKNENILFDSISVLARKNNIKRLKNYMKNNKCYIYDDNNNK 1785

OY 52 ECYYNVVDGEELDOEKFYVDENFTENYL 79
 Db 1786 K-KKNNKNVENQEKREYVINKIFVINFI 1812

RESULT 11

T32731
 PAR interacting protein - rat

C:Species: Rattus norvegicus (Norway rat)
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T32731

R:Comte, P.A.; Ossipow, V.; Schibler, U.
 submitted to the EMBL Data Library, January 1997

A:Description: Isolation of PIP, a 160 kDa nucleolar protein that interacts with the actin
 A:Reference number: 221213
 A:Accession: T32731
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1277 <COM>
 A:Cross-references: EMBL:U83590; NID:g2253210; PID:g2253211; PIDN:AAB62878.1

Query Match 13.0%; Score 74.5; DB 2; Length 1277;
 Best Local Similarity 30.4%; Pred. No. 53;
 Matches 31; Conservative 9; Mismatches 37; Indels 25; Gaps 5;

QY 1 EREDCTWYANRKYDPD-KSPKSSDDECKKTCFTECYIVFEDVNEKCYNNVD 59
 DB 630 EDED-----NVVDTDEKOLKHEGSDS-----EDSKNSDVSDD 670
 QY 60 GEELDQ--KFVVDENFTENYLTDEGKDGAGNAAGTGDSEDE 99
 DB 671 GESEEDRDNDKVDPEFRQOLM---EVLQAGNALGSEEE 709

RESULT 12
 D70465
 hypothetical protein aq_1923 - Aquifex aeolicus
 C:Species: Aquifex aeolicus
 C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 05-Nov-1999
 C:Accession: D70465
 R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O'V.
 Nature 392, 353-358, 1998
 A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
 A:Reference number: A70300; M0ID:98196666; PMID:9537320

A:Accession: D70465
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-296 <AO>
 A:Cross-references: GB:AE000763; NID:g2984178; PIDN:AAC07719.1; PID:g2984187; GB:AE00065
 A:Experimental source: strain VFS
 C:Genetics:
 A:Gene: aq_1923
 C:Superfamily: Aquifex aeolicus hypothetical protein aq_1923

Query Match 12.9%; Score 74; DB 2; Length 296;
 Best Local Similarity 24.2%; Pred. No. 14;
 Matches 22; Conservative 18; Mismatches 21; Indels 30; Gaps 4;

QY 9 FYANRKT-----FDKSKSSDDECKKTCFTECYIVFEDVNEKCY 54
 DB 18 YFAYDKYMDNKAKEQVEYFLDKTLKSG-----KGYKYVDYKPIGELIKDY 67
 QY 55 YNVVDGEELDQKFPVVDENFTENYLTDEGK 85
 DB 68 YRDNNGEFPKTEELIIRK-----LSFTECK 92

RESULT 13
 B70236
 hypothetical protein BBH09 - Lyme disease spirochete Plasmid H/1p28-3
 C:Species: Borrelia burgdorferi (Lyme disease spirochete)
 C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
 C:Accession: B70236
 R:Fraser, C.M.; Castles, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
 son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
 Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
 Nature 390, 580-586, 1997
 A:Authors: Smith, H.O.; Venter, J.C.
 A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.
 A:Reference number: A70100; M0ID:9805945; PMID:9403685
 A:Accession: B70236
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1278 <KLE>
 A:Cross-references: GB:AE000784; NID:g2690041; PIDN:AAC66000.1; PID:g2690056; TIGR:BBH09

A:Experimental source: strain B31
 C:Genetics:
 A:Genome: plasmid

Query Match 12.9%; Score 74; DB 2; Length 1278;
 Best Local Similarity 25.0%; Pred. No. 59;
 Matches 18; Conservative 17; Mismatches 25; Indels 12; Gaps 3;

QY 10 YANRKYDPEKSPKSSDDECKKTCFTECYIVFEDVNEKCYNNVD-----VDGEELD 64
 DB 181 YGEKRIEFD--FKRIKEE-----YKEQEMFLFSYLIRKERYLKTSLNISVEKQIS 233
 QY 65 QEKVEVDENFTE 76
 DB 234 KEKEIQKTRE 245

RESULT 14
 S17231
 dynein beta heavy chain, ciliary - sea urchin (Anthracidaris crassispina)
 N:Contains: dynein ATPase (EC 3.6.4.2)
 C:Species: Anthracidaris crassispina
 C:Date: 30-Sep-1991 #sequence_revision 02-May-1994 #text_change 19-Apr-2002
 C:Accession: S17231; PS0415
 R:Ogawa, K.
 Nature 352, 643-645, 1991
 A:Title: Four ATP-binding sites in the midregion of the beta heavy chain of dynein.
 A:Reference number: S17231; M0ID:91326104; PMID:1830928
 A:Accession: S17231
 A:Molecule type: mRNA
 A:Residues: 1-4466 <OGA>
 A:Cross-references: GB:D01021; NID:g217202; PIDN:BA00827.1; PID:g217203
 R:Ogawa, K.
 Proc. Jpn. Acad. B Phys. Biol. Sci. 67, 27-31, 1991
 A:Title: ATP-binding site in dynein beta-heavy chain: identification by molecular clo

A:Reference number: PS0415
 A:Accession: PS0415
 A:Molecule type: mRNA
 A:Residues: 764-1001, 'APG', 1005-2036, 'VPSSVET' <OGZ>
 C:Superfamily: dynein heavy chain, ciliary
 C:Keywords: ATP; heterotrimer; hydrolase; microtubule binding; nucleotide binding;
 F:154-161/Region: nucleotide-binding motif A (P-loop)
 F:185-185/Region: nucleotide-binding motif A (P-loop)
 F:2133-2140/Region: nucleotide-binding motif A (P-loop)
 F:2460-2467/Region: nucleotide-binding motif A (P-loop)
 F:2805-2812/Region: nucleotide-binding motif A (P-loop)
 F:160/Binding site: ATP (Lys) #status predicted
 F:1858/Binding site: ATP (Lys) #status predicted
 F:2139/Binding site: ATP (Lys) #status predicted
 F:2466/Binding site: ATP (Lys) #status predicted
 F:2811/Binding site: ATP (Lys) #status predicted

Query Match 12.9%; Score 74; DB 1; Length 4466;
 Best Local Similarity 25.9%; Pred. No. 2,1e+02;
 Matches 28; Conservative 14; Mismatches 28; Indels 38; Gaps 6;

QY 1 EREDCTWYANRKY-----TFDKSKSSDDECKKTCFTECYIVFEDVNEKCYNNVD-----CK-----KTC 35
 DB 460 EEEQKAKVFTPEYDGLDPTCOEFLDEEEKVFFVLDRLGSLTQGFDDCGLEAA 519
 QY 36 FKTECY-----IVFEDVNEKCYNNVD-----DGEELDQKFPVVDEN 73
 DB 520 FKMLDCTGYLDRPVRINDF-----ECKYPIVLMYQELDQSKSEIYDEH 564

RESULT 15
 T21428
 hypothetical protein p26P2.6 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T21428
 R:Ainscough, R.
 submitted to the EMBL Data Library, November 1996

Wed Feb 26 15:59:05 2003

us-09-936-737a-2.rpr

Page 5

```

A:Reference number: 219420
A:Accession: T21428
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-643 <MIL>
A:Cross-References: EMBL:Z81514; PIDDN:CAB04192.1; GSPDB:GND0023; CESP:F26F2.6
C:Genetics:
A:Gene: CESP:F26F2.6
A:Map position: 5
A:Introns: 18/3; 116/3; 142/1; 265/3; 306/3; 373/3; 403/2; 425/2; 479/3; 544/3; 573/2; 6
  Query Match      12.7%; Score 73; DB 2; Length 643;
  Best Local Similarity 26.6%; Pred. No. 37;
  Matches 29; Conservative 14; Mismatches 32; Indels 34; Gaps 6;

OY      7 WTFY-----ANKRY--DFOKSPKKS-----SDLDECKKTFKTEICYAFED 47
      1:1:      11      11      1:1:      1:1:      1:1:      1:1:      1:1:
Db      155 WSEFPPYDGDANCASVTMVODYAKGPETVAKPRDAOISNMDECHMCEEEEOCIAYILD 214
      .:1:1:      11      11      1:1:      1:1:      1:1:      1:1:      1:1:

OY      48 TVNREC--YYNVNDEEELDQE-----KFWVDN--FTENYLYLD 81
      .:1:1:      11      11      1:1:      1:1:      1:1:      1:1:      1:1:
Db      215 SDAKCAVMYSSDDGLTFPMKSSADSGKRLVIAKMAVDETTCLETPPQILD 263

```

Search completed: February 25, 2003, 02:33:20
Job time : 16 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 25, 2003, 02:26:35 ; Search time 13 Seconds

(without alignments)
328.620 Million cell updates/sec

Title: US-09-936-737A-2

Perfect score: 573
Sequence: 1 EREDCWTFYANRKYTFDK.....GKDAAGNAGTGDESDDEVDED 103

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 segs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	85.5	14.9	147	1	LAPP_HAEOR
2	80.5	14.0	412	1	FBK4_SPOFR
3	78	13.6	797	1	VG48_HSVSA
4	75.5	13.2	419	1	ICEC_MOUSE
5	75.5	13.2	491	1	P2C1_CAEEL
6	75	13.1	360	1	Y0R4_CAEEL
7	74	12.9	296	1	Y073_MOUSE
8	74	12.9	4466	1	DYHC_ANTCR
9	73	12.7	1405	1	DPOA_SCHPO
10	71	12.4	782	1	YNY2_YEAST
11	70.5	12.3	78	1	MOH1_CANPG
12	70.5	12.3	146	1	PA21_MOUSE
13	70.5	12.3	156	1	CB20_HUMAN
14	70.5	12.3	748	1	Y875_METUA
15	70	12.2	1058	1	SI85_YEAST
16	69.5	12.1	146	1	PA21_PIG
17	68.5	12.0	610	1	CALG_HUMAN
18	68	11.9	146	1	PA21_MOUSE
19	68	11.9	240	1	NER_HY2D2
20	68	11.9	4466	1	DYHC_TRIGR
21	67.5	11.8	277	1	PS12_ARATH
22	67.5	11.8	602	1	SPR8_YEAST
23	67.5	11.8	321	1	SPR8_YEAST
24	67	11.7	146	1	PA21_RAT
25	67	11.7	327	1	RREP_PIRYV
26	67	11.7	650	1	CC45_YEAST
27	67	11.7	2410	1	MOK1_SCHPO
28	66.5	11.6	78	1	MOH2_SCHPO
29	66.5	11.6	125	1	FABL_ANOPU
30	66.5	11.6	132	1	PA21_HORSE
31	66.5	11.6	405	1	LHX1_BRARE
32	66.5	11.6	591	1	CALX_RAT
33	66.5	11.6	733	1	Y008_YEAST

34	66.5	11.6	1033	1	S190_YEAST	P36123 saccharomyc
35	66	11.5	148	1	CB20_XENIA	P52299 xenopus lae
36	66	11.5	1647	1	YDEL_SCHPO	O10435 schizosacch
37	65.5	11.4	165	1	ADFL_CAEEL	Q07750 caenorhabdi
38	65.5	11.4	428	1	RF4_KLULA	P09806 kluyveromyc
39	65.5	11.4	455	1	YN06_YEAST	P40161 saccharomyc
40	65.5	11.4	725	1	RIR1_MYCTU	P50640 mycobacteri
41	65.5	11.4	746	1	GYP7_YEAST	P48365 saccharomyc
42	65.5	11.4	793	1	NETB_DROME	O24568 drosophila
43	65	11.3	1124	1	TCF8_HUMAN	P37275 homo sapien
44	65	11.3	1281	1	TRPS_HUMAN	Q9uh17 homo sapien
45	65	11.3	1616	1	VTT1_CAEEL	P5155 caenorhabdi

ALIGNMENTS

RESULT 1	LAPP_HAEOR	STANDARD:	PRT:	147 AA.
ID	001747			
AC	01-FEB-1994 (Rel. 28, Created)			
DT	01-FEB-1994 (Rel. 28, Last sequence update)			
DE	01-FEB-1994 (Rel. 28, Last annotation update)			
GN	Anti-platelet protein precursor.			
OS	Haemateria officinalis (Mexican leech).			
OC	Eukaryota; Metazoa; Annelida; Clitellata; Hirudindae; Hirudinea;			
OC	Rhynchobdellidae; Glossiphoniidae; Haemateria.			
OX	NCBI_TaxID=6410;			
RP	[1]			
RC	SEQUENCE FROM N.A., AND SEQUENCE OF 60-91 AND 123-139.			
RA	TISSUE=Salivary gland.			
RA	MEDLINE=92202247; PubMed=1551898;			
RT	Keller P.M., Schultz L.D., Condra C., Karczewski J., Connolly T.M.;			
RT	"An inhibitor of collagen-stimulated platelet activation from the			
RT	salivary glands of the Haemateria officinalis leech. II. Cloning of			
RT	the cDNA and expression."			
RL	J. Biol. Chem. 267:6899-6904(1992).			
CC	-1- FUNCTION: AN INHIBITOR OF COLLAGEN-STIMULATED PLATELET			
CC	AGGREGATION, DENSE GRANULE RELEASE AND SEROTONIN RELEASE.			
CC	-1- SUBCELLULAR LOCATION: Secreted.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@lsb-sib.ch).			
CC	-----			
DR	EMBL; M81489; AAA29194.1; -.			
KW	SIGNAL.			
FT	SIGNAL 1 21			
FT	CHAIN 22 147			
FT	CHAIN ANTI-PLATELET PROTEIN.			
SO	SEQUENCE 147 AA; 15908 MW; 75A5511374AAE4ZE CRC64;			
Query Match	14.9%; Score 85.5; DB 1; Length 147;			
Best local Similarity	33.9%; Pred. No. 0.2;			
Matches 19; Conservative 10; Mismatches 24; Indels 3; Gaps 1;				
OY	4 EDCWTFYANRKYTFDKSFKRSKSDLDCKKTCFTEY---CYVFEDTVNKECYN 56			
DB	65 EDCWTKRRGKWLPLDNLTKTEFTSVDECRKNCESAVPSCTLIQINTETNECRN 120			
RESULT 2	FKB4_SPOFR	STANDARD:	PRT:	412 AA.
ID	026486			
AC	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	01-NOV-1999 (Rel. 38, Last annotation update)			

DE 46 kDa FK506-binding nuclear protein (Peptidyl-prolyl cis-trans
 DE isomerase) (PPIase) (EC 5.2.1.8).
 OS Spodoptera frugiperda (Fall armyworm).
 CC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 CC Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;
 CC Ditrysia; Noctuidae; Noctuidae; Amphipyrinae; Spodoptera.
 NCBI_TaxID=7108;
 [1]
 RP SEQUENCE FROM N.A. PubMed=7527037;
 RX MEDLINE=95074110; Fernandes-Alnemri T., Pomeroy K., Robertson N.M.,
 RA Alnemri E.S., Fernandes-Alnemri T., Pomeroy K., Robertson N.M.,
 RA Duley K., Dubois G.C., Litwack G.;
 RT "FKBP6, a novel Src tyrosine kinase inhibitor that forms a
 RT protein-kinase complex.";
 RT J. Biol. Chem. 269:30828-30834(1994).
 CC -1- FUNCTION: BINDS TO, AND IS INHIBITED BY FK506 AND RAPAMYCIN. BINDS
 CC DOUBLE-STRANDED DNA IN VITRO. PPIASES ACCELERATE PROTEIN FOLDING.
 CC -1- CATALYTIC ACTIVITY: CIS-TRANS ISOMERIZATION OF PROLINE IMIDIC
 CC PEPTIDE BONDS IN OLIGOPEPTIDES.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- PTM: PHOSPHORYLATED BY A NUCLEAR KINASE IN THE PRESENCE OF MG2+
 CC AND ATP.
 CC -1- SIMILARITY: BELONGS TO THE FKBP-TYPE PPIASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U15038; AAA58962.1; -.
 DR HSSP: P20071; FRKJ.
 DR InterPro: IPR001179; FKBP_PPIase.
 DR PIRam: PF00234; FKBP_1.
 DR PROSITE: PS00453; FKBP_PPIASE_1; FALSE_NEG.
 DR PROSITE: PS00454; FKBP_PPIASE_2; 1.
 DR PROSITE: PS50059; FKBP_PPIASE_3; 1.
 KW Isomerase; Rotamase; Nuclear protein; DNA-binding; Phosphorylation.
 FT DOMAIN 91 112 ASP/GLU-RICH (HIGHLY ACIDIC).
 FT DOMAIN 120 145 LYS-RICH (BASIC).
 FT DOMAIN 152 216 ASP/GLU-RICH (HIGHLY ACIDIC).
 FT DOMAIN 219 302 LYS-RICH (BASIC).
 FT DOMAIN 324 412 PPIASE, FKBP-TYPE.
 SQ SEQUENCE 412 AA; 45810 MW; F2A69159AEFFFE22 CRC64;
 Query Match 14.0%; Score 80.5; DB 1; Length 412;
 Best Local Similarity 30.6%; Pred. No. 1.7;
 Matches 30; Conservative 12; Mismatches 29; Indels 27; Gaps 5;
 QY 11 ANRKRTDFDKSPFKS-----SDLECKKTKCTKCYIVFEDYVYVVDGEELDQ 65
 DB 133 ANKRAKPKPKKAKGNSAPAESDSDDD-----EDLQK-----FLDGEDIDT 174
 QY 66 EKFDVDEFTENYLTDCGCKDAGNAAGTGDSEDEVD 103
 DB 175 DE--NDESFKN--TSAEGDSDSEDEDEDEDD 208
 RESULT 3
 VG48_HSVSA STANDARD: PRT; 797 AA.
 AC 001033;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 01-APR-1993 (Rel. 25, Last annotation update)
 DE Hypothetical gene 48 protein.
 GN 48 OR EDLFS.
 OS Herpesvirus saimiri (strain 11).
 CC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 CC Gammaherpesvirinae; Rhadinovirus.
 NCBI_TaxID=10383;
 CC

RN [1]
 RP SEQUENCE FROM N.A. PubMed=1321287;
 RX MEDLINE=92333688; Albrecht J.-C., Nicholas J., Biller D., Cameron K.R., Bisinger B.,
 RA Albrecht J.-C., Nicholas J., Biller D., Cameron K.R., Bisinger B.,
 RA Newman C., Wiltmann S., Craxton M.A., Coleman H., Fleckenstein B.,
 RA Honess R.W.;
 RT "Primary structure of the herpesvirus saimiri genome.";
 RT J. Virol. 66:5047-5058(1992).
 CC -1- SIMILARITY: TO EBV BBRF2.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X64346; CAA45671.1; -.
 DR PIR: A36811; A36811.
 KW Hypothetical protein.
 SQ SEQUENCE 797 AA; 88667 MW; 9E294234AD850E23 CRC64;
 Query Match 13.6%; Score 78; DB 1; Length 797;
 Best Local Similarity 32.3%; Pred. No. 5.7;
 Matches 31; Conservative 12; Mismatches 33; Indels 20; Gaps 5;
 QY 11 ANRKRTDFDKSPFKSDDLECKKTKCTKCYIVFEDYVYVVDGEELDQ 66
 DB 382 ANRK--EYKTIIDSDRDRDRDEYELE-----NEE--YNRDEEDEGEDEDE 427
 QY 67 KFDVDEFTENYLTDCGCKDAGNAAGTGDSEDEVD 102
 DB 428 KDRKEGDEGDEGDEGDEGEF--DEGDEGDEGE 461
 RESULT 4
 ICEC_MOUSE STANDARD: PRT; 419 AA.
 AC 008736;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Caspase-12 precursor (EC 3.4.22.-).
 GN CASP12.
 OS Mus musculus (Mouse).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C3H/An.
 RX MEDLINE=97190206; PubMed=9038361;
 RA van de Craen M., Vandanaebele P., Declercq W., van den Brande I.,
 RA van Loo G., Molemans F., Schotte P., van Crielinge W., Beyaert R.,
 RT Fiers W.;
 RT "Characterization of seven murine caspase family members.";
 RT FEBS Lett. 403:61-69(1997).
 CC -1- FUNCTION: INVOLVED IN THE ACTIVATION CASCADE OF CASPASES
 CC RESPONSIBLE FOR APOPTOSIS EXECUTION (BY SIMILARITY).
 CC -1- SUBUNIT: HETERODIMER OF TWO SUBUNITS (POTENTIAL).
 CC -1- TISSUE SPECIFICITY: MAINLY EXPRESSED IN SKELETAL MUSCLE AND LUNG.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C14.
 CC -1- SIMILARITY: CONTAINS 1 CARD DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: Y13090; CAAT3532.1; -
 DR HSSP: P29466; 1ICE.
 DR MEROPS: C14.013; -
 DR MGD: MGI:J31922; Casp12.
 DR InterPro: IPR001315; CARD.
 DR InterPro: IPR002398; ICE.
 DR InterPro: IPR002138; ICE_P10.
 DR InterPro: IPR001309; ICE_P20.
 DR Pfam: PF00655; ICE_P10; 1.
 DR Pfam: PF00656; ICE_P20; 1.
 DR PRINTS: PR00376; TLICENZIME.
 DR SMART: SM00114; CARD; 1.
 DR SMART: SM00115; CAS; 1.
 DR PROSITE: PS01029; CARD; 1.
 DR PROSITE: PS01121; CASPASE_CYS; 1.
 DR PROSITE: PS01121; CASPASE_HIS; 1.
 DR PROSITE: PS0207; CASPASE_P10; 1.
 DR PROSITE: PS0208; CASPASE_P20; 1.
 KW Hydrolyase; Thiol protease; Apoptosis; Zymogen.
 FT PROPEP 1 ? ?
 FT CHAIN 1 ? ?
 FT PROPEP 2 ? ?
 FT CHAIN 2 ? ?
 FT ACT_SITE 250 250
 FT ACT_SITE 298 298
 FT ACT_SITE BY SIMILARITY.
 SQ SEQUENCE 419 AA; 47853 MW; B94B0FED16B1CB40 CRC64;

Query Match 13.2%; Score 75.5; DB 1; Length 419;
 Best Local Similarity 35.6%; Pred. No. 5;
 Matches 26; Conservative 14; Mismatches 14; Indels 19; Gaps 6;

QY 44 VFEDYVKECYNVVDEEL---DOEKTVVD--ENFTNYL--IDCEK-DAGNAAGTG 94
 ID VFEDYVKECYNVVDEEL---DOEKTVVD--ENFTNYL--IDCEK-DAGNAAGTG 94
 Db 26 VFEDYVKECYNVVDEEL---DOEKTVVD--ENFTNYL--IDCEK-DAGNAAGTG 94
 QY 95 DE-----SDEVD 101
 ID DE-----SDEVD 101
 Db 82 EQLSLQPSNDEDD 94

RESULT 5
 P2CL_CAEEL STANDARD; PRT; 491 AA.
 ID P2CL_CAEEL STANDARD; PRT; 491 AA.

AC P49595;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Probable protein phosphatase 2C F42G9.1 (EC 3.1.3.16) (P2C).
 GN F42G9.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Pelodermidae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Bristol N2;
 RA Tatch A.; Waterston R.;
 RA Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
 CC -1 CATALYTIC ACTIVITY: A phosphoprotein + H(2)O - a protein +
 CC phosphate.
 CC -1 SIMILARITY: BELONGS TO THE P2C FAMILY.
 CC -1 SIMILARITY: BELONGS TO THE P2C FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL; 000051; AAA91358.1; -

DR HSSP: P35813; 1A60.
 DR WormPep: F42G9.1; CE07231.
 DR InterPro: IPR001932; P2C-1like.
 DR InterPro: IPR000222; P2C.
 DR Pfam: PF00481; P2C; 1.
 DR SMART: SM00331; P2C-SIG; 1.
 DR SMART: SM00332; P2C; 1.
 DR PROSITE: PS01032; P2C; 1.
 KW Hypothetical protein; Hydrolyase; Magnesium; Manganese.
 FT METAL 37 37
 FT METAL 38 38
 FT METAL 57 57
 FT METAL 428 428
 FT METAL 477 477
 FT METAL MANGANESE 1 (BY SIMILARITY).
 FT METAL MANGANESE 1 AND 2 (BY SIMILARITY).
 FT METAL MANGANESE 2 (BY SIMILARITY).
 FT METAL MANGANESE 2 (BY SIMILARITY).
 SQ SEQUENCE 491 AA; 53141 MW; F110D12E3A3953D6 CRC64;

Query Match 13.2%; Score 75.5; DB 1; Length 491;
 Best Local Similarity 24.5%; Pred. No. 5.9;
 Matches 23; Conservative 18; Mismatches 28; Indels 25; Gaps 4;

QY 10 YANRKYTF--DKSPKSSDLDECKTKCTEYCYFEDTVNKECYNVVDEELDOEK 67
 Db 67 YANRKYTF--DKSPKSSDLDECKTKCTEYCYFEDTVNKECYNVVDEELDOEK 67
 QY 68 FVVDENFTENYLLDCEGKDGNAAGTGDSDEVD 101
 Db 113 DIDDEG-----KPKRAGEADSEDEADRID 137

RESULT 6
 YOF_A_CAEEL STANDARD; PRT; 360 AA.
 ID YOF_A_CAEEL STANDARD; PRT; 360 AA.

AC Q09337;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Hypothetical 39.7 kDa protein ZK1290.10 in chromosome II precursor.
 GN ZK1290.10.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Pelodermidae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Bristol N2;
 RA Tatch A.;
 RA Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL; 021308; AAB93316.1; -

DR WormPep: ZK1290.10; CE02080.
 DR InterPro: IPR003582; SHKT.
 DR SMART: SM00254; SHKT; 1.
 KW Hypothetical protein; Signal.
 FT SIGNAL 1 21
 FT CHAIN 22 360
 FT SIGNAL HYPOTHETICAL PROTEIN ZK1290.10.
 SQ SEQUENCE 360 AA; 39669 MW; 18751B6BC062DAF7 CRC64;

Query Match 13.1%; Score 75; DB 1; Length 360;
 Best Local Similarity 33.3%; Pred. No. 4.8;
 Matches 23; Conservative 9; Mismatches 23; Indels 14; Gaps 3;

QY 46 EDTVYVKECYNVVDEELDOEKTVVD--ENFTNYL--IDCEGKDGNAAGTG 94
 Db 224 EDTVYVKECYNVVDEELDOEKTVVD--ENFTNYL--IDCEGKDGNAAGTG 94

MEDLINE=91238692; PubMed=2034212;
Damagnez V., Tillitt J., de Recondo A.-M., Baldacci G.,

SQ	SEQUENCE	1405 AA;	159348 MM;	06082A3BAD6347CZ CRC64;
	Query Match	12.7%;	Score 73;	DB 1; Length 1405;
	Best Local Similarity	32.3%;	Pred. NO. 30;	
	Matches	21; Conservative	10; Mismatches	18; Indels 16; Gaps
OY	45 FEDIVNKECYINVNDG-----ELDGEKFEVDEN--FTENYLTDCEKGADAGNAAGT 94	: : : : : :	:	:
Dd	31 YESKENELIDVNSEETKRIVQRLEDDFVVDDNGAGYVDN-----GYDEWDOSHYH 84	: : : : : :	:	:
OY	95 DESDE 99	: : : : : :	:	:
Dd	85 DEDDE 89	: : : : : :	:	:

RESULT	10
YNV2 YEAST	*
ID YNV2 YEAST	STANDARD; PRT; 782 AA.
P40157;	
DT 01-FEB-1995 (Rel. 31, Created)	
DT 01-FEB-1995 (Rel. 31, Last sequence update)	
DT 16-OCT-2001 (Rel. 40, Last annotation update)	
DE Hypothetical 98.6 kDa protein in PEK17-MER1 intergenic region.	
GN YNL212W OR N1327.	
OS Saccharomyces cerevisiae (Baker's yeast).	
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;	
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.	
XX NCBI_TaxID=4932;	
RN [1]	
RP SEQUENCE FROM N.A.	
RC STRAIN=S288C / FYA1679;	
RA MEDLINE=95282517; PubMed=7762305;	
EX Coster F., van Dyck L., Joniaux J.-L., Purnelle B., Goffeau A.;	
RT "The sequence of a 13.5 kb DNA segment from the left arm of yeast	
RT chromosome XIV reveals MER1: RAPI, a new putative serine/threonine phosphatase	
RT replication complex and a new putative serine/threonine phosphatase	
RT gene".	
RL Yeast 11.85-91(1995).	
CC -1 SIMILARITY: TO C.CARDUNCULUS PROTEIN CYP604 AND S.POMBE	
CC SPEC1685.1AC.	

CC This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -	
CC the European Bioinformatics Institute. There are no restrictions on its	
CC use by non-profit institutions as long as its content is in no way	
CC modified and this statement is not removed. Usage by and for commercial	
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/	
CC or send an email to license@isb-sib.ch).	

CC EMBL: X78898; CAA55496.1; -	
DR EMBL: Z71488; CAA96114.1; -	
KW SCD; S0005156; YNL212W.	
KW Hypothetical protein.	
DW SCQUENCE 782 AA; 88845 MW; CCBAIF4CE32BF97C CRC64;	
SQ	
Query Match	12.4%; Score 71; DB 1; Length 782;
Best Local Similarity	26.5%; Pred. NO. 25;
Matches 22; Conservative 16; Mismatches 39; Indels 6; Gaps 2	
OY 24 KSSLDSECKRKTCTKCYTYIVEDIVNKECYYNAVVDGE---LQEKFYVDENTFNYLT 80	
: :: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :	
DB 315 KFRLLDCICIFRFRAWVKCLMM--TLNKKETWTDVEKEKYDYLDSVSPLEKOPDDLIHI 371	
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :	
OY 81 DCEGKDAGNAAGTCGDSDEVDED 103	
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :	
DB 372 DRSNEEDRKESSESDSED 394	
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :	
RESULT 11	
MOHL_CANDG	STANDARD; PRT; 78 AA.
ID MOHL_CANDG	
AC P81034;	
DT 15-JUL-1998 (Rel. 36, Created)	

DT 15-JUL-1998 (Rel. 36, last sequence update)
 DE 15-JUN-2002 (Rel. 41, last annotation update)
 DE Mandibular organ-inhibiting hormone-1 (MOIH-1).
 OS Cancer pagurus (Rock crab).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Crustacea;
 OC Malacostraca; Eumalacostraca; Eucarida; Decapoda; Pleocyemata;
 OC Brachyura; Eubrachyura; Cancroidea; Cancridae; Cancer.
 NCBI_TaxID=6755;
 RN [1]
 RP SEQUENCE.
 RC TISSUE=Sinus gland;
 RX MEDLINE=96278659; PubMed=8662685;
 RA Mainwright G., Webster S.G., Wilkinson M.C., Chung J.S., Rees H.H.;
 RT "Structure and significance of mandibular organ-inhibiting hormone in
 the crab, Cancer pagurus. Involvement in molting and reproduction."
 RL J. Biol. Chem. 271:12749-12754 (1996).
 CC -1- FUNCTION: REPRESSSES THE SYNTHESIS OF METHYL FARNESATE, THE
 CC -1- PRECURSOR OF INSECT JUVENILE HORMONE III IN THE MANDIBULAR ORGAN.
 CC -1- TISSUE SPECIFICITY: PRODUCED BY THE MEDULLA TERMINALIS X-ORGAN IN
 CC THE EUSTALKS AND TRANSPORTED TO THE SINUS GLAND WHERE IT IS
 CC STORED AND RELEASED.
 CC -1- SIMILARITY: BELONGS TO THE ARTHROPOD CHH/MH/GH/VIH FAMILY OF
 CC HORMONES.
 DR InterPro: IPR001166; CHH_MH_GH.
 DR Pfam: PF01147; Crust_neurochem; 1.
 DR PRINTS: PR00550; HYPERGLICEMIC.
 DR PROSITE: PS01250; CHH_MH_GH; 1.
 DR KW Neuropeptide; Hormone.
 FT DISULFID 7 44 BY SIMILARITY.
 FT DISULFID 24 40 BY SIMILARITY.
 FT DISULFID 27 53 BY SIMILARITY.
 SQ SEQUENCE 78 AA; 9241 MW; 8DB338A39058A65D CRC64;
 Query Match 12.3%; Score 70.5; DB 1; Length 78;
 Best Local Similarity 33.9%; Pred. No. 2.5;
 Matches 20; Conservative 5; Mismatches 23; Indels 11; Gaps 3;
 Oy 5 DCMFFYNRRY-----TDFKSPFKSSDDECKTCF-KTEV--CYTFEDTVNKE 52
 Db 6 DCONFIGNRAMYKRVKWDICDCANIPRKDGLNNRSCNCFYTFELWCIDATENTRNKE 64
 RESULT 12
 PA21_CANFA STANDARD; PRT; 146 AA.
 AC P06596;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 15-JUN-1988 (Rel. 06, last sequence update)
 DE 15-JUN-2002 (Rel. 41, last annotation update)
 DE Phospholipase A2 precursor (EC 3.1.1.4) (phosphatidylcholine 2-
 acylhydrolase) (Group IB phospholipase A2).
 GN PLA2G1B.
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 NCBI_TaxID=9615;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86223862; PubMed=3754861;
 RA Ohara O., Tamaki M., Nakamura E., Tsuruta Y., Fujii Y., Shin M.,
 RA Teraoka H., Okamoto M.;
 RT "Dog and rat pancreatic phospholipases A2: complete amino acid
 sequences deduced from complementary DNAs."
 RL J. Biochem. 99:733-739 (1986).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90306027; PubMed=2142076;
 RA Kerfelec B., Latorge K.S., Vasiloudes P., Puigserver A., Scheele G.A.;
 RT "Isolation and sequence of the canine pancreatic phospholipase A2
 gene."
 RL Eur. J. Biochem. 190:299-304 (1990).
 RN [3]

RP SEQUENCE FROM N.A.
 RX MEDLINE=87175472; PubMed=3562437;
 RA Kerfelec B., Latorge K.S., Puigserver A., Scheele G.A.;
 RT "Primary structures of canine pancreatic lipase and phospholipase A2
 messenger RNAs."
 RL Pancreas 1:430-437 (1986).
 CC -1- FUNCTION: PA2 catalyzes the calcium-dependent hydrolysis of the
 CC -1- 2-acyl groups in 3-sn-phosphoglycerides.
 CC -1- CATALYTIC ACTIVITY: Phosphatidylcholine + H₂O = 1-
 CC -1- acylglycerophosphocholine + a fatty acid anion.
 CC -1- COFACTOR: Binds 1 calcium ion per subunit (By similarity).
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE A2 FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 DR EMBL: D00035; BAA00023.1;
 DR EMBL: M35301; AAA30883.1;
 DR PIR: A24392; PSDG.
 DR PIR: S11316; S11316.
 DR HSSP: P00592; 4P2P.
 DR InterPro: IPR001211; PhospholipaseA2.
 DR Pfam: PF00068; Phoslip; 1.
 DR PRINTS: PR00389; PHPLIPASEA2.
 DR PRODOM: PD000303; PhospholipaseA2.
 DR SMART: SM00085; PA2c; 1.
 DR PROSITE: PS00118; PA2_HIS; 1.
 DR PROSITE: PS00119; PA2_ASP; 1.
 DR KW Hydrolyase; Lipid degradation; Calcium; Pancreas; Signal.
 FT SIGNAL 1 15
 FT PROPEP 1 15
 FT CHAIN 23 146
 FT ACT_SITE 23 146
 FT ACT_SITE 70 70
 FT ACT_SITE 121 121
 FT DISULFID 33 99
 FT DISULFID 49 146
 FT DISULFID 51 67
 FT DISULFID 66 127
 FT DISULFID 73 120
 FT DISULFID 83 113
 FT DISULFID 106 118
 FT CA_BIND 50 50
 FT CA_BIND 52 52
 FT CA_BIND 54 54
 FT CA_BIND 71 71
 SQ SEQUENCE 146 AA; 16235 MW; F6258ED9527F3692 CRC64;
 Query Match 12.3%; Score 70.5; DB 1; Length 146;
 Best Local Similarity 25.4%; Pred. No. 4.9;
 Matches 15; Conservative 11; Mismatches 18; Indels 15; Gaps 1;
 Oy 28 LDECKTCFETKYCYIVEEDVNEKCYINVDGELDQEFVVDENFTNYLTDEGKD 86
 Db 60 VDELDKCOHTDHCT-----SEAKLDSCKFLDNDNYTYKYSYSGSGS 103
 RESULT 13
 CB20_HUMAN STANDARD; PRT; 156 AA.
 AC P52298; Q14924;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, last sequence update)
 DT 15-JUN-2002 (Rel. 41, last annotation update)
 DE 20 kDa nuclear cap binding protein (NCBP 20 kDa subunit) (CBP20) (NCBP
 interacting protein 1) (NIP1).
 GN NCBP2 OR CBP20.
 OS Homo sapiens (human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 [1]
 RP SEQUENCE FROM N.A. AND SEQUENCE OF 9-25 AND 113-145.
 RX MEDLINE=95379956; PubMed=7651522;
 RA Izaurralde E., Lewis J., Gambrell C., Jarmolowski A.,
 RA McGuigan C., Malta A.W.;
 RT "A cap-binding protein complex mediating U snRNA export.";
 RL Nature 376:709-712(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Cervical carcinoma;
 RX MEDLINE=96017765; PubMed=7478990;
 RA Kataoka N., Ohno M., Mada I., Shimura Y.;
 RT "Identification of the factors that interact with NCBP, an 80 kDa
 RT nuclear cap binding protein.";
 RL Nucleic Acids Res. 23:3638-3641(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Cervix;
 RA Strausberg R.;
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: INVOLVED IN MEDIATING U SNRNA EXPORT FROM THE NUCLEUS.
 CC BINDS TO 5/CAPPED MRNA.
 CC -1- SUBUNIT: THE CAP-BINDING PROTEIN (CBC) COMPLEX IS AN HETERODIMER
 CC OF CBP80 AND CBP20.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: X84157; CAAS8962.1; -
 DR EMBL: D59253; BA09599.1; -
 DR EMBL: BC001255; AA01255.1; -
 DR HSSP: P11940; ICVJ.
 DR Genew: HGNC:7659; NCBP2.
 DR MIM: 605133; -
 DR InterPro: IPR000504; RNA_rec_mot.
 DR Pfam: PF00076; rrm; 1.
 DR SMART: SM00360; RRM; 1.
 DR PROSITE: PS0102; RRM; 1.
 DR PROSITE: PS00030; RRM_RNP_1; FALSE_NEG.
 KW Nuclear protein; RNA-binding.
 FT DOMAIN 40 118 RNA-BINDING (RRM).
 FT CONFLICT 97 97 A -> S (IN REF. 2).
 SQ SEQUENCE 156 AA; 18001 MW; B6C94F3182A2CC3D CRC64;
 Query Match 12.3%; Score 70.5; DB 1; Length 156;
 Best Local Similarity 21.8%; Pred. No. 5.2;
 Matches 26; Conservative 20; Mismatches 50; Indels 23; Gaps 6;
 QY 1 EERE-----DCWTFYANRKYTDPEKSS-----FKKSSD-----LDECKKTCFTECYIV 44
 DB 30 EDEKLLKSKCTLYVGMISYTTTEQYELFFSSGDIKIMLDMKKTYA--CGCFWE 87
 QY 45 FEDTVNKECYNVVDEGEELDOEKFFVDENFTENYLTDCGKDGAGNAAGTDESDEVD 103
 DB 88 YYSRADAENMRYINGTRLDLR--IIRTDWDAFK--EGROYGKRSRGQYRDEYRQD 141
 RESULT 14
 Y875_METUA STANDARD; PRT; 748 AA.
 AC 058285;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein MJ0875.
 GN MJ0875.
 OS Methanococcus jannaschli.
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
 OC Methanocaldococcaceae; Methanocaldococcus.
 OX NCBI_TaxID=2190;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=8688087;
 RA Bolt C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
 RA Overbeek A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.L.,
 RA Scott J.L., Kirschness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Uitterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
 RT jannaschli.";
 RL Science 273:1058-1073(1996).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: U67531; AAB98882.1; -
 DR TIGR: MJ0875; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 748 AA; 90731 MW; 5515ED3B541562NA CRC64;
 Query Match 12.3%; Score 70.5; DB 1; Length 748;
 Best Local Similarity 24.4%; Pred. No. 27;
 Matches 29; Conservative 16; Mismatches 39; Indels 35; Gaps 6;
 QY 1 EEREDCWTFY-----ANKRYTDFPKSFKSSDDECKKTCFK-TECYIVFED--- 47
 DB 144 KEFSIDWYFIDKIYKYKLSIKRKEGNDKAIELTE--KQYLAESYNNKPDKNY 200
 QY 48 -----TVNKECYNVVDEGEELDOEKFFVDENFTENYLTDCGKDGAGNAAGTDESDEVD 102
 DB 201 KKAELTFNHFYNNIM-----AKFESERKFE-----AAEYKKSGDITKEIDE 244
 RESULT 15
 S185_YEAST STANDARD; PRT; 1058 AA.
 ID S185_YEAST
 AC P40856;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE S184-associated protein SAPI85.
 GN SAPI85 OR YJ098W OR J0840
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=96220458; PubMed=8649382;
 RX Luke M.M., della Seta F., di Como C.J., Sugimoto H., Kobayashi R.,
 RA Arndt K.T.;
 RT "The SAP, a new family of proteins, associate and function positively
 RT with the S184 phosphatase.";
 RL Mol. Cell. Biol. 16:2744-2755(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288c;

```

RX MEDLINE-95176706; PubMed-7871887;
RA Miosga T., Boles E., Schaaf-Gerstenschlaeger I., Schmitt S.,
RA Zimmermann F.K.;
RT "Sequence and function analysis of a 9.74 kb fragment of
RT Saccharomyces cerevisiae chromosome X including the BCK1 gene.";
RU Yeast 10:1481-1488(1994).
RM [3]
RP SEQUENCE OF 1-853 FROM N.A.
RC STRAIN-S288c / FY1679;
RX MEDLINE-96090136; PubMed-7483851;
RA Rasmussen S.W.;
RT "A 37.5 kb region of yeast chromosome X includes the SWE1, MER2, GSH1
RT and CSD3 genes, a TCP-1-related gene, an open reading frame similar
RT to the DAL80 gene, and a tRNA(Arg).";
RU Yeast 11:873-883(1995).
CC -! FUNCTION: ASSOCIATES WITH THE SIT4 PHOSPHATASE IN A CELL CYCLE
CC DEPENDENT MANNER. MAY BE DIRECTLY OR INDIRECTLY INVOLVED IN
CC SIT4-DEPENDENT FUNCTIONS IN BUDDING AND IN NORMAL G1 CYCLIN
CC EXPRESSION.
CC -! PTM: HYPERPHOSPHORYLATED IN THE ABSENCE OF SIT-4.
CC -! SIMILARITY: BELONGS TO THE SAPS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X77923; CAA54882.1; -
DR EMBL: X85021; CAA59396.1; -
DR EMBL: Z49373; CAA89392.1; -
DR SGD: S0003634; SAP185.
KW Phosphorylation; Cell cycle.
SQ SEQUENCE 1058 AA; 121402 MW; DD44DD53DD87438 CRC64;

Query Match 12.2%; Score 70; DB 1; Length 1058;
Best Local Similarity 20.3%; Pred. No. 43;
Matches 25; Conservative 27; Mismatches 45; Indels 26; Gaps 5;

QY 2 ERECCWTFYANRRKYTFDFKSFKSSDLDECKKTCFTEY-----CYVFE-- 46
DB 427 DRDITYIGLYVMESEHMAFNKI--LTEKKIPLLOTSTYGTIEPLGFERRFKICELIAEL 484
QY 47 -----DRVNKECYNVVGEELDEKEKYVDENFE-NYLDCGKXGAGNAGGDSDEY 100
DB 485 HCSNMTLNLNPSAYDIVREDAERERLEFNSQNYVDSNDRSELEKNEEDN---TGDADDEV 541
QY 101 DED 103
DB 542 EDD 544

```

Search completed: February 25, 2003, 02:31:32
 Job time : 15 secs

1	78.5	13.7	116	16	08X41	08X41 clostridium
2	78.5	13.7	509	2	09X91	09X91 streptococ
3	78.5	13.7	550	5	09U24	09U24 caenorrhadi
4	78.5	13.7	907	5	09X79	09X79 caenorrhadi
5	78	13.6	682	16	09K85	09K85 bacillus ha
6	78	13.6	3933	5	09V23	09V23 plasmodium
7	77	13.4	368	5	09VH8	09VH8 dirosophila
8	77	13.4	374	5	08SY5	08SY5 dirosophila
9	77	13.4	264	5	003650	003650 trypanosoma
10	75.5	13.2	945	5	020246	020246 caenorrhadi
11	75	13.1	3844	5	094648	094648 plasmodium
12	74.5	13.0	606	10	09FIW8	09FIW8 aridiopsist
13	74.5	13.0	1277	11	035821	035821 rattus norv
14	74	12.9	596	10	09FK41	09FK41 aridiopsist
15	74	12.9	1278	16	050667	050667 borrelia bu
16	73.5	12.8	156	11	09CQ49	09CQ49 mus musculu

17	73.5	12.8	1154	5	Q9V6S9	Q9V6S9 drosophila
18	73	12.7	643	5	Q9XV53	Q9XV53 caenorhabdi
19	73	12.7	2543	5	P90649	P90649 paramesli
20	72.5	12.7	564	5	O01627	O01627 caenorhabdi
21	72.5	12.7	2533	5	O27183	O27183 paramesli
22	72.5	12.7	2533	5	P90589	P90589 paramesli
23	72	12.6	1303	10	Q9SKS0	Q9SKS0 arabidopsis
24	71.5	12.5	320	5	O96214	O96214 plasmodium
25	71.5	12.5	540	16	O8XHG9	O8XHG9 clostridium
26	71.5	12.5	594	16	O9P6E1	O9P6E1 campylobact
27	71	12.4	551	4	Q9NX93	Q9NX93 homo sapien
28	71	12.4	589	4	Q9HA35	Q9HA35 homo sapien
29	71	12.4	851	4	Q9H501	Q9H501 homo sapien
30	70.5	12.3	112	5	Q9U5Y6	Q9U5Y6 cancer pagu
31	70.5	12.3	256	10	Q9CA59	Q9CA59 arabidopsis
32	70.5	12.3	296	11	Q9D5G6	Q9D5G6 mus musculu
33	70.5	12.3	367	5	Q9VIR9	Q9VIR9 drosophila
34	70.5	12.3	600	16	O8RH66	O8RH66 fusobacteri
35	70.5	12.3	786	10	O9C7E7	O9C7E7 arabidopsis
36	70.5	12.3	2472	5	O8FTM5	O8FTM5 dictyosteli
37	70	12.2	292	12	O6S3Z9	O6S3Z9 african swi
38	70	12.2	328	5	O22144	O22144 caenorhabdi
39	70	12.2	521	5	O94718	O94718 paramesli
40	70	12.2	669	5	O8SXS8	O8SXS8 drosophila
41	70	12.2	703	10	Q9LKR9	Q9LKR9 arabidopsis
42	70	12.2	793	10	Q9SPU5	Q9SPU5 zea mays (m
43	70	12.2	1167	16	O9ZLG7	O9ZLG7 helicobacteri
44	70	12.2	1255	16	O97IT6	O97IT6 clostridium
45	70	12.2	1481	5	Q9NEX0	Q9NEX0 caenorhabdi

ALIGNMENTS

RESULT 1

ID	PRELIMINARY;	PRT;	116 AA.
OBX141			
AC	OBX141		
DT	01-MAR-2002 (TREMBLrel. 20, Created)		
DT	01-MAR-2002 (TREMBLrel. 20, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	Hypothetical protein CPE1920.		
GN	CPE1920.		
OS	Clostridium perfringens.		
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;		
OC	Clostridiales; Clostridiaceae; Clostridium.		
OX	NCBI_Taxid=1502;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=13 / TYPE A;		
RX	PubMed=11792842;		
RA	Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,		
RA	Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.,		
RT	"Complete genome sequence of Clostridium perfringens, an anaerobic		
RT	flesh-eater."		
RL	Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).		
DR	EMBL: AP063192; BAB1626.1; -.		
DR	InterPro: IPR001440; TPR.		
DR	Pfam: PF00515; TPR; 3.		
DR	SMART: SM00028; TPR; 3.		
KW	Hypothetical protein; Complete proteome.		
SC	SEQUENCE 116 AA; 13590 MW; 9DB5BD847B3A7312 CRC64;		

Query Match	13.7%;	Score 78.5;	DB 16;	Length 116;
Best Local Similarity	22.7%;	Pred. No. 1.9;		
Matches	22;	Conservative	11;	Mismatches 27;
				Indels 37;
				Gaps 2

```

QY 10 YANKRTYDPKSPKSSDLECK-----KTC 35
      | : | : | : | : |
      | : | : | : | : |
Db 11 YNTFNYSDAIINYKKALDNECKCHSYNAGCYI:KTKQYEKALEMITKALBELYQDSKFE 70
      | : | : | : | : |
QY 36 FKTEYCYIVFEDYVNEKCYIYN---VDGELDQKEFV 69
      | : | : | : | : |

```


OC Bacillaceae: Bacillus.
 NCBI_TaxID=86665;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C-125 / JCM 9153;
 RX MEDLINE=20512582; PubMed=11058132;
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
 RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
 RA Horikoshi K.,
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
 RT halodurans and genomic sequence comparison with Bacillus subtilis.",
 RT Nucleic Acids Res. 28:4317-4331(2000).
 CC -1- SIMILARITY: TO PROKARYOTE SENSORY TRANSDUCTION PROTEINS.
 DB EMBL: AF001517, BAB06689.1; .
 DR HSSP: 056310; 1B30.
 DR InterPro: IPR003594; ATPbind_ATPase.
 DR InterPro: IPR004358; Bact_sens_pr_C.
 DR InterPro: IPR002345; Chem.
 DR InterPro: IPR004105; H-kinase_dtm.
 DR InterPro: IPR004359; HIS_KIN_sig.
 DR InterPro: IPR002570; Hpt.
 DR Pfam: PF01584; Chem; 1.
 DR Pfam: PF02895; H-kinase_dtm; 1.
 DR Pfam: PF02318; HATPase_c; 1.
 DR Pfam: PF01627; Hpt; 1.
 DR PRINTS: PR00344; BCTRLSENSOR.
 DR PRODOM: PD003142; Hpt; 1.
 DR SMART: SM00260; Chem; 1.
 DR SMART: SM00387; HATPase_c; 1.
 DR SMART: SM00073; Hpt; 1.
 DR Kinase: Phosphorylation; Sensory transduction; Transferase;
 KW Complete proteome.
 SQ SEQUENCE 682 AA; 75272 MW; 46B6B1EE5BEF7276 CRC64;

Query Match 13.6%; Score 78; DB 16; Length 682;
 Best Local Similarity 32.6%; Pred. No. 15;
 Matches 26; Conservative 13; Mismatches 31; Indels 14; Gaps 5;

QY 10 YANRKYTDKDFKSSD-----LDECKTKCTKTECYVFE--DTVNEKCYNNYVG 60
 DB 156 YDEFEMVLDOSFEGYAYOIEVTLDE--KTLKAAVFEVLEBOVG-EVTKSPSPA 212
 QY 61 EELDOEKVVDENFENYLTDCDGD 86
 DB 213 EELDEK--DERFLVTLITVDEGE 236

RESULT 6

O97239 PRELIMINARY; PRT; 3933 AA.

AC O97239
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Hypothetical 467.9 kDa protein.
 GN PFC0245C, MAL3P2.18.
 OS Plasmodium falciparum (isolate 3D7).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
 OX NCBI_TaxID=36329;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=3D7;
 RX MEDLINE=99376085; PubMed=10448855;
 RA Bowman S., Lawson D., Basham D., Brown D., Chillingworth T.,
 RA Churcher C.M., Craig A., Davies R.M., Devlin K., Feltwell T.,
 RA Gentles S., Gilliam R., Hamlin N., Harris D., Holtroyd S., Hornsby T.,
 RA Hurrecks P., Jagski K., Jaseal B., Kyes S., McLean J., Moule S.,
 RA Murrell K., Murphy L., Oliver K., Quail M.A., Rajandream M.-A.,
 RA Rutter S., Skellon J., Squares R., Squares S., Sulston J.E.,
 RA Whitehead S., Woodward J.R., Newbold C., Barrell B.G.;
 RT "The complete nucleotide sequence of chromosome 3 of Plasmodium
 RT falciparum.",
 RL Nature 400:532-538(1999).

DR EMBL: AL034558; CAB39005.1; .
 DR InterPro: IPR002048; EF-hand.
 DR PROSITE: PS00018; EF_HAND; UNKNOWN_1.
 KW Hypothetical protein.
 SQ SEQUENCE 3933 AA; 467876 MW; 5144A4604EE36933 CRC64;

Query Match 13.6%; Score 78; DB 5; Length 3933;
 Best Local Similarity 25.0%; Pred. No. 1e+02;
 Matches 23; Conservative 21; Mismatches 24; Indels 24; Gaps 5;

QY 36 FKTEYC-YIVF-----EDTVNKECY--NVVDGEILD-QEKF-----VVD 71
 DB 1139 FETNNCEVITIFLFYVHLLIEKELKSCFELKNIINNCCEMKNKFEVFCFLHITR 1198
 QY 72 ENFENYLTDCGKAGNAAAGTGDSEVDYD 103
 DB 1199 INFNKSLLKNYKIKRAGDITDDDDDDDDDD 1230

RESULT 7

O9VPH8 PRELIMINARY; PRT; 368 AA.

AC O9VPH8
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE CG14853 protein.
 GN CG14853.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abrial J.F., Agbayani A., An H.-J., Andrews-Frankoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu U., Beasley E.M.,
 RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA De Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fouts R.C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegam C.,
 RA Jalali M., Kalush F., Kapen G.H., Ke Z., Kennison U.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Matrei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spadling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weisenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

CLASS	NUMBER	PERCENT	COMMENTS
CARBONID	6/9	6/9	
N-LINKED (GLCNAC. . .)			(POT

53 CYYNVDGEELDQEK-VVDENITE--NYLIDCEGKDAAGNAAGIGDESDER

Yamada K., Banh J., Chang C.H., Chang E., Dale J.M., Golds RA

SEQUENCE FROM N. A.
PB
RN [2]

RP SEQUENCE FROM N.A.
RA Yamada K., Banh J., Chang C.H., Chang E., Dale J.M., Goldsmith A.D.
RA Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M., Wu H.C.,
RA

RA Yamamura Y., Yu G., Yu S., Bowser L., Caroincl P., Chen H., Cheuk R.,
 RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
 RA Kawai J., Kim C., Lam B., Lin J., Meyers M.C., Miranda M.,
 RA Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,
 RA Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
 RA Theologis A.,
 RT "Full Length cDNA of gene At5g39780 (GI:15242471)."
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB016876; BAB1383.1; -
 DR EMBL; AY070102; AAL49839.1; -
 KW Hypothetical protein
 SQ SEQUENCE 606 AA; 71043 MW; 5F02E244E59FC15D CRC64;

Query Match 13.0%; Score 74.5; DB 10; Length 606;
 Best Local Similarity 27.7%; Pred. No. 29;
 Matches 31; Conservative 19; Mismatches 37; Indels 25; Gaps 7;
 QY 1 EEREPCW--TFYANRKYTFD---KSKKSSDLDECKKTCFTECYIVFEDTVNKECY 55
 Db 103 EKTEDLGVSFHGDKKTKEDYSVSFKK-----KTRFLTE-----EDFLESDS-- 147
 QY 56 NVVDGEEL---DQEKVDENFTNYLFDCEGKAGNAGTDESDEVED 103
 Db 148 DFVDSQTFSTNDEDFLSDPAETSLKKGQNRKSDN--SGSGSDESEEE 198

RESULT 13
 035821 PRELIMINARY; PRT; 1277 AA.
 AC 035821;
 DT 01-JUN-1998 (TREMBlrel. 05, Created)
 DT 01-JUN-1998 (TREMBlrel. 05, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE PAR interacting protein.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91084854; PubMed=2261643;
 RA Martin J., Schibler U.;
 RT "Expression of the liver-enriched transcriptional activator protein
 RT DBP follows a stringent circadian rhythm.";
 RL Cell 63:1257-1266 (1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Conte P.A., Ossipow V., Schibler U.;
 RT "Isolation of PIP, a 160 kDa nuclear protein that interacts with the
 RT activation domain of PAR transcription factors.";
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U83590; AAB62878.1; -
 SQ SEQUENCE 1277 AA; 144675 MW; 43140A2973178792 CRC64;

Query Match 13.0%; Score 74.5; DB 11; Length 1277;
 Best Local Similarity 30.4%; Pred. No. 66;
 Matches 31; Conservative 9; Mismatches 37; Indels 25; Gaps 5;
 QY 1 EEREPCWTFYANRKYTFD--KSKKSSDLDECKKTCFTECYIVFEDTVNKECY 59
 Db 630 EDEED-----NVVDTDEKQLKHGSDASD-----EDSKNSSEVDSDSD 670
 QY 60 GEELDQF--KRVVDENFTNYLFDCEGKAGNAGTDESDE 99
 Db 671 GESEEDRDKVDYDPCFQOLM--EVLQAGNALGGESEEE 709

RESULT 14
 09FK41 PRELIMINARY; PRT; 596 AA.
 AC 09FK41;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Similarity to unknown protein.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; Rosidae;
 OC Eucosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RX MEDLINE=98403884; PubMed=9734815;
 RA Krcani H., Nakamura Y., Sato S., Asamizu E., Kaneko T., Miyajima N.,
 RA Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. VI.
 RT Physically assigned p1 and TAC clones.";
 RL DNA Res. 5:203-216 (1998).
 DR EMBL; AB012248; BAB09450.1; -
 DR InterPro; IPR001798; Kelch.
 DR Pfam; PF01344; Kelch; 6.
 SQ SEQUENCE 596 AA; 68483 MW; 8865D3CAB873C8D0 CRC64;

Query Match 12.9%; Score 74; DB 10; Length 596;
 Best Local Similarity 24.2%; Pred. No. 32;
 Matches 29; Conservative 23; Mismatches 40; Indels 28; Gaps 5;
 QY 12 NRRYTFDSEK-----KSSDLDECKKTCFTECYIVFEDTV-----NKEC 53
 Db 339 NRRWSDSKRALQCAKTRLPQVSSEYVKKPCGRINSCNVGKDTLYIGGMEIKREV 398
 QY 54 YYN--VVGGEELDQKRVFVDENFTN--NYLTDCEG-----KDAGNAGTDESDEVED 103
 Db 399 TLDDLYSLNLSKLDKWCIIPTTEWEVSDDEGDEDDDESDSESGNSESDDDDD 458

RESULT 15
 050667 PRELIMINARY; PRT; 1278 AA.
 AC 050667;
 DT 01-JUN-1998 (TREMBlrel. 06, Created)
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Hypothetical protein BBH09.
 GN BBH09.
 OS Borrelia burgdorferi (Lyme disease spirochete).
 OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
 OX NCBI_TaxID=139;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 35210 / B31;
 RX MEDLINE=98065943; PubMed=9403685;
 RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
 RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
 RA Dougeny J.B., Tomb J.F., Fleischmann R.D., Richardson S., Hanson M.,
 RA Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Wainman J.,
 RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
 RA Uterback T., Watney L., McDonald L., Attlich P., Bowman C.,
 RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
 RA Smith H.O., Venter J.C.;
 RT "Genomic sequence of a Lyme disease spirochete, Borrelia
 RT burgdorferi.";
 RL Nature 390:580-586 (1997).
 DR EMBL; AE000784; AAC66000.1; -
 DR TIGR; BBH09; -
 DR InterPro; IPR002296; N12M6_mtfase.
 DR InterPro; IPR002052; N6_mtfase.
 DR PRINTS; PR00507; N12M6_mtfase.
 DR PROSITE; PS00092; N6_mtfase; UNKNOWN 1.
 KW Hypothetical protein; Plasmid; Complete proteome.
 SQ SEQUENCE 1278 AA; 150871 MW; 8A0B22DF166565C0 CRC64;

Query Match 12.9%; Score 74; DB 16; Length 1278;

Wed Feb 26 15:59:07 2003

us-09-936-737a-2.rsp

Page 7

Best Local Similarity 25.0%; Pred. NO. 74;
Matches 18; Conservative 17; Mismatches 25; Indels 12; Gaps 3;

```

QY      10 YANRKYTDDEKSEKSSDLDECKKTKCTEYCIIVFEDTVNKEGYNV-----VDGEELD 64
      | ::||| | : : | : : || | : : | : : | : : | : : | : : | : : |
Db     181 YGEKRYIEFD--FSKIKEKEE-----YKEQEWELFYSYLIRKERYLKTSNISIVEKEQIS 233

```

```
QY 65 QEKFVDENFTE 76
    :||:::|
Db 234 KEKEIIQKTLRE 245
```

Search completed: February 25, 2003, 02:32:52
Job time : 27 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 25, 2003, 02:33:51; Search time 1946 Seconds

(without alignments)
857.212 Million cell updates/sec

Title: US-09-936-737a-2

Perfect score: 573

Sequence: 1 EREDCCWTFYANKRYTDFDK.....GKDAAGNAAGTDESDEVED 103

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+pn.model -DEV=xlh
-Q/cgn2.1/USPTO.spool/US09936737/runat_14022003_100733_2299/app_query.fasta_1.263
-DB=Est -OPMT=fastap -SUFFIX=st -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=human40.cdi -LIST=45
-LOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pro -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09936737@cgn2.1_1349@runat_14022003_100733_2299 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-MARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST.*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	84.5	14.7	504	10 BE581359	BE581359 Kq49904.Y
2	83	14.5	704	17 A0643182	A0643182 RRC193-EC
3	82	14.3	462	12 BG141291	BG141291 EST481733
4	82	14.3	470	12 BG137845	BG137845 EST478287
5	82	14.3	484	12 BG141201	BG141201 EST481643
6	82	14.3	493	12 BG141314	BG141314 EST481756
7	82	14.3	496	12 BG139798	BG139798 EST480240
8	82	14.3	529	12 BG136045	BG136045 EST476487
9	82	14.3	541	10 AW737851	AW737851 EST339278
10	82	14.3	548	12 BG139905	BG139905 EST480347
11	82	14.3	561	12 BG136094	BG136094 EST476536
12	82	14.3	561	12 BG138707	BG138707 EST479149
13	82	14.3	561	12 BG139936	BG139936 EST480378
14	82	14.3	561	12 BG140418	BG140418 EST480860
15	82	14.3	561	12 BG140421	BG140421 EST480863
16	82	14.3	567	12 BG139637	BG139637 EST479955
17	82	14.3	597	12 BG136116	BG136116 EST476558
18	82	14.3	600	10 AW738722	AW738722 EST340149
19	82	14.3	620	10 AW218016	AW218016 EST266731
20	82	14.3	648	12 BG140407	BG140407 EST480849
21	82	14.3	652	12 BG136420	BG136420 EST476862
22	82	14.3	663	10 AW738268	AW738268 EST339695
23	82	14.3	672	12 BG137992	BG137992 EST478434
24	82	14.3	687	12 BG136406	BG136406 EST476848
25	82	14.3	693	10 AW738284	AW738284 EST339711
26	82	14.3	693	13 BI933236	BI933236 EST33125
27	82	14.3	706	13 BI933320	BI933320 EST553209
28	82	14.3	723	13 BI933312	BI933312 EST553201
29	80.5	14.0	499	12 BG138008	BG138008 EST478450
30	80	14.0	594	12 BG138662	BG138662 EST479104
31	79.5	13.9	656	12 BF708357	BF708357 A1200_LE
32	79.5	13.9	1058	17 CENS0505	AL34161 Tetradon
33	79	13.8	470	12 BG138190	BG138190 EST478632
34	78	13.6	566	13 BJ101346	BJ101346 BJ101346
35	78	13.6	947	17 A2676398	A2676398 ENRJO51TF
36	77	13.4	410	17 AA549977	AA549977 1056m3 gm
37	77	13.4	663	17 BH316991	BH316991 CH230-120
38	76	13.3	508	17 A0651585	A0651585 Sheared D
39	76	13.3	721	9 AL726177	AL726177 AL726177
40	76	13.3	1051	10 BE259556	BE259556 601154720
41	75.5	13.2	473	12 BG086881	BG086881 H3131G02
42	75.5	13.2	478	13 BI174761	BI174761 OSTF053A8
43	75.5	13.2	526	17 A0701970	A0701970 HS.5396.B
44	75.5	13.2	640	13 BJ122797	BJ122797 BJ122797
45	75.5	13.2	643	10 BB645658	BB645658 BB645658

ALIGNMENTS

RESULT 1
LOCUS BE581359
DEFINITION BE581359 504 bp mRNA linear EST 09-MAY-2001
sequence.
ACCESSION BE581359
VERSION BE581359.1 GI:9832301
KEYWORDS EST.
ORGANISM Strongyloides stercoralis.
Strongyloides stercoralis.
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditidae;
Panagrolaimoidae; Strongyloidae; Strongyloides.
REFERENCE 1 (bases 1 to 504)
McCarter, O., Clifton, S., Chiapelli, B., Pape, D., Martin, J., Wylie, T.,
Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, Y.

TITLE
JOURNAL
COMMENT
Gibbons, M., Rutter, E., Bennett, J., Franklin, C., Tsagaris, R., Ronko, L., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe, M., Allen, M., Person, B., Swaller, T., Harrey, N., Schurk, R., John, S., Shih, T., Jackson, Y., Cardenas, M., McCan, K., Waterston, R., and Wilson, R.
The Washington Univ. Nematode EST Project, 1999
Unpublished (1999)
Contact: McCarter JP
The Washington Univ. Nematode EST Project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
The library was constructed by Dr. Thomas Nutman and colleagues of NIAID, NIH (nutman@nih.gov). DNA Sequencing by: Washington University Genome Sequencing Center St. Louis.
High quality sequence stop: 414.

FEATURES

source
1. 504
Location/Qualifiers
/organism="Strongyloides stercoralis"
/strain="Rhabditiform larvae obtained from gerbils"
/db_xref="taxon:6248"
/clone_lib="TBN95TM-SSR"
/lab_host="XL-1 Blue MRF" (Stratagene)
/note="Vector: Lambda Uni-ZAP XR (Stratagene); Site 1: EcoRI; Site 2: XhoI; mRNA was purified from 2 x 10E3 rhabditiform larvae which had been isolated from gerbils experimentally infected with larvae originally isolated from experimentally infected dogs. cDNA was constructed and, using adaptors, was cloned unidirectionally into the vector from the EcoRI site to the XhoI site. The library has an unamplified titer of 1 x 10E5 pfu/ml and an amplified, undiluted titer of 9 x 10E11 pfu/ml. The average insert size of the unamplified library is 675 bp (range: 100-1700)."
BASE COUNT 200 a 62 c 59 g 183 t
ORIGIN

Alignment Scores:

Pred. No.: 0.754 Length: 504
Score: 84.50 Matches: 19
Percent Similarity: 54.90% Conservative: 9
Best Local Similarity: 37.25% Mismatches: 19
Query Match: 14.75% Indels: 4
DB: 10 Gaps: 2

US-09-936-737a-2 (1-103) x BE581359 (1-504)

QY 9 PheTYrAlaAsnArgLysTYrThr-AspPheAspLysSerPheLysLysSerSerAspLeu 28
:::||||| ||||| ||| ||||| ::| |||||
Db 319 TGTATACACAAATAGCAAAAGACAAAGATATATCCATCTAGAAATGCAATCTAGATT 378
QY 28 uAspGluCysLysLysThrCysPheLysThrGluTYrCysTYr---lLeValPheGluAs 47
| ||||| ||| ||||| ::||| ||| ::|||
Db 379 AAAAGAAATGATTAACAAAGTATGCAATCTTTTGTATTCATTAATCTAGATGA 438
QY 47 pThrValAsnLysGluCys-----TYrTYr 55
: ||||| : ||| : |||||
Db 439 AACATATTAATATCGTTGTCTTTTATTTAT 469
RESULT 2
LOCUS A0643182 704 bp DNA linear GSS 08-JUL-1999
DEFINITION RPI93-ECORI-6L8.TJ RPI93-ECORI Trypanosoma brucei genomic clone
ACCESSION A0643182
VERSION A0643182.1 GI:5119892
KEYWORDS GSS.
SOURCE Trypanosoma brucei.
ORGANISM Trypanosoma brucei
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.

REFERENCE

1 (bases 1 to 704)
AUTHORS El-Sayed, N., Zhao, S., Zhao, H., Gill, S., Suh, E., Malek, J., Fujii, C., Gerrard, C., Leech, V., de Jong, P., Ullu, E., Melville, S., Donelson, J., Fraser, C., and Adams, M.
TITLE Use of BAC end sequences from Trypanosoma brucei GUTat 10.1 RPI93 Library for gene discovery and sequence-ready map construction
JOURNAL Unpublished (1999)
COMMENT Other GSSs: RPI93-ECORI-6L8.TJ
Contact: Najib M. El-Sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: nelsayed@igf.org

FEATURES

source
1. 704
Location/Qualifiers
/organism="Trypanosoma brucei"
/strain="TREU927/4 GUTat 10.1"
/db_xref="taxon:5691"
/clone_lib="RPI93-ECORI-6L8"
/note="Vector: pBAC3.6; Site 1: Eco RI; Site 2: Eco RI; constructed for The Institute for Genomic Research by Bohui Zhao in Pieter de Jong's laboratory (Roswell Park Cancer Institute, Buffalo, NY). Briefly, Trypanosoma brucei TREU927/4 GUTat 10.1 agarose embedded DNA was partially digested with a combination of Eco RI and Eco RI methylase (RPI93-ECORI segment) or Dpn II (RPI93-DpnII segment). High molecular weight fragments were ligated in pBAC3.6 vector digested with Eco RI or Bam HI, respectively. The average insert size is 141 Kb. Total coverage (both segments): > 90 x the haploid non-mitochondrial genome."

BASE COUNT 290 a 92 c 147 g 175 t
ORIGIN

Alignment Scores:

Pred. No.: 1.77 Length: 704
Score: 83.00 Matches: 26
Percent Similarity: 40.43% Conservative: 12
Best Local Similarity: 27.66% Mismatches: 24
Query Match: 14.49% Indels: 32
DB: 17 Gaps: 3

US-09-936-737a-2 (1-103) x A0643182 (1-704)

QY 11 AlaAsnAlaGlyTYrThr-AspPheAspLysSerPheLysLysSerSerAspLeuAspLeu 30
||| ||| ||| ::||| ||| |||
Db 70 GCAACCCCGAAGACAAACAGAAACAACTTCGAAAGAAAGAAAGAGATATA 129
QY 31 CysLysLysThrCysPheLysThrGluTYrCysTYrllLeValPheGluAspThrValAsn 50
| |||||
Db 130 TGCAAA----- 135
QY 51 LysGluCysTYrTYrAsnValValAspLysGluGluLeuAspGluGluLysPheValVal 70
| |||||
Db 136 -----GATGCG-----TCCAAGATGGGCTTCAATGG 162
QY 71 AspGluAsnPheThrGluAsnTYrLeuThrAspCysGluGlyLysAsp-----Ala 87
| ||||| ||| : ||||| ||| : |||
Db 163 GATGAGAAATGCACATGCAAAAGCAAAAGGAGCTGCAAACTTAAGAAAGAAAGAAAGACA 222
QY 88 GlyAsnAlaIaGlyThrGlyAspGluSerAspGluValAsp 101
| ||||| ||||| ||||| ||||| |||||
Db 223 AACACAGCAGCAGACAGACAGACAGATACAGAGGATGAT 264
RESULT 3

LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	COMMENT	FEATURES
BGI41291	BGI41291	462 bp	mrna	linear	EST 31-JAN-2001						
	EST481733 wild tomato pollen Lycopersicon pennellii			CDNA clone							
	CLPP20L16 5' sequence, mRNA sequence.										
BGI41291	BGI41291	462 bp	mrna	linear	EST 31-JAN-2001						
	EST.										
	Lycopersicon pennellii.										
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;										
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;										
	Asteridae; euasterids I; Solanales; Solanaceae; Solanum;										
	Lycopersicon.										
	1 (bases 1 to 462)										
	van der Hoeven,R., Bezzerides,J., Sun,H., Cho,J., Uterback,T.,										
	Hansen,C., Reuning,C. and Tanksley,S.										
	Generation of ESTs from wild tomato (L. pennellii) pollen										
	Unpublished (2001)										
	Contact: CUGI										
	Clemson University Genomics Institute										
	Clemson University										
	100 Jordan Hall, Clemson, SC 29634, USA										
	Email: http://www.genome.clemson.edu/orders/index.html .										
	Location/Qualifiers										
	1..462										
	/organism="Lycopersicon pennellii"										
	/cultivar="TA56"										
	/db_xref="taxon:28526"										
	/clone="CLPP20L16"										
	/clone_1lb="wild tomato pollen"										
	/tissue_type="pollen"										
	/dev_stage="pollen collected from open flowers"										
	/lab_host="SOLR"										
	/note="Vector: pBluescript SK(-); Site_1: EcoRI, Site_2:										
	XhoI; Pollen was collected from open flowers from										
	L.pennellii TA56, and stored at -80 C until library										
	construction."										
BASE COUNT	175 a	85 c	68 g	134 t							
ORIGIN											
Alignment Scores:											
Pred. No.:	1.42										
Score:	82.00										
Best Similarity:	48.68%										
Best Local Similarity:	34.21%										
Query Match:	14.31%										
DB:	12										
	Gaps:										
	4										
US-09-936-737A-2 (1-103) x BGI41291 (1-462)											
QY	14	LysTyrThrAspPhe-----AspLysSerPheLysLysSerSerAspLeu	28								
DB	238	AAAGCAACAAGATTTTATCAAAAATGTAATTCATATAAAGATTGGCTTCACAGTCACACATA	297								

KEYWORDS	EST.
SOURCE	Lycopersicon pennellii.
ORGANISM	Lycopersicon pennellii
	Eukaryote: Vitridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae: euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE	1 (bases 1 to 470)
AUTHORS	van der Hoeven, R., Bezzerides, J., Sun, H., Cho, J., Utterback, T., Hansen, C., Rönning, C. and Tanksley, S.
TITLE	Generation of ESTs from wild tomato (L. pennellii) pollen
JOURNAL	Unpublished (2001)
COMMENT	Contact: CUGI Clemson University Genomics Institute Clemson University 100 Jordan Hall, Clemson, SC 29634, USA Email: http://www.genome.clemson.edu/orders/index.html .
FEATURES	Location/Qualifiers
source	1..470 /organism="Lycopersicon pennellii" /cultivar="TA56" /db_xref="taxon:28526" /clone="cLPP7D15" /clone_lib="wild tomato pollen" /tissue_type="pollen" /dev_stage="pollen collected from open flowers" /lab_host="SOLR" /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; Pollen was collected from open flowers from L.pennellii TA56, and stored at -80 C until library construction."
BASE COUNT	177 a 86 c 70 g 137 t
ORIGIN	
Alignment Scores:	
Pred. No.:	1.45 Length: 470
Score:	82.00 Matches: 26
Percent Similarity:	48.68% Conservative: 11
Best Local Similarity:	34.21% Mismatches: 27
Query Match:	14.31% Indels: 12
DB:	12 Gaps: 4
US-09-936-737A-2 (1-103) x BG137845 (1-470)	
QY	14 LysTyrThrAsphe-----AspLysSerPheLysSerSerAspLeu 28
Db	238 AAGACACACAGATTTTTCACAAAATGTAATTGTAATAAGATGGCTTCAGTCACACCATTA 297
QY	29 AspGluCysLysLysThrCysPheLysThrGluTyrCysTyrIleValPheLusprtr 48
Db	298 GATCATATATAGAAAGATGTTTAAACA-----TGCAGAGAGGTGTATAGATGCA 351
QY	49 ValAsn-----LysGluCysTyrTyrAsnValValAspGlyGluLusAspGlnu 66
Db	352 ATGATTCATATGAAAAAGCACAACAAGAGTT-----AAAGAACTAATATTATAC 402
QY	67 LysPheValValAspGluAsnpherhGluAsnTyrLeuThrAspCys 82
Db	403 AAAGCAAAATATGATATTGTGTCATCATATACCTTGACACTTGT 450
RESULT 5	
LOCUS	BG141201 484 bp mRNA linear EST 31-JAN-2001
DEFINITION	ESTAB1643 wild tomato pollen Lycopersicon pennellii cDNA clone
ACCESSION	BG141201
VERSION	BG141201.1 GI:12641390
KEYWORDS	EST.
SOURCE	Lycopersicon pennellii.
ORGANISM	Lycopersicon pennellii
	Eukaryote: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum;

FEATURES	source	Location/Qualifiers
CONTACT	CUGI	Contact: CUGI
INSTITUTION	Clemson University Genomics Institute	Clemson University
ADDRESS	100 Jordan Hall, Clemson, SC 29634, USA	
EMAIL	http://www.genome.clemson.edu/orders/index.html.	
FEATURES		
source	1..493	
	/organism="Lycopersicon pennellii"	
	/cultivar="TA56"	
	/db_xref="taxon:28526"	
	/clone="cLPP20P16"	
	/clone_lib="wild tomato pollen"	
	/tissue_type="pollen"	
	/dev_stage="pollen collected from open flowers"	
	/lab_host="SODR"	
	/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; Pollen was collected from open flowers from L.pennellii TA56, and stored at -80 C until library construction."	
BASE COUNT	185 a	91 c 74 g 142 t 1 others
ORIGIN		
Alignment Scores:		
Pred. No.:	1.53	Length: 493
Score:	82.00	Matches: 26
Percent Similarity:	48.68%	Conservative: 11
Best Local Similarity:	34.21%	Mismatches: 27
Query Match:	14.31%	Indels: 12
DB:	12	Gaps: 4
US-09-936-737A-2 (1-103) x BG141314 (1-493)		
QY	14 LysTYrThrAspPhe-----AspLysSerPheLysSerSerAspLeu 28	
DB	238 AAGACACACAGATTATATCAAAAGATGTAATGATTAAGATTGGCTTCAAGTCAACCATATA 297	
QY	29 AspGluCysLysLysThrCysPheLysThrGluTYrLysCysTYrLeuValPheGluAspThr 48	
DB	298 GATCATTTATTAAGAAAGATTGTTTAAACA-----TCAAGAGAGCTGATGAAGATGCA 351	
QY	49 ValAsn-----LysGluCysTYrTYrAsnValValAspGlyGluGluLeuAspGlnGlu 66	
DB	352 ATTGATTCATGATAAAAAAGGCCACACAAAGATGTT-----AAAGAACTAAATTATTAC 402	
QY	67 LysPheValValAspGluAsnPheThrGluAsnTYrLeuTYrLeuThrAspCys 82	
DB	403 AAAGCAATAATNGATATTGCTGCATCATATACACTTGACACTTGT 450	
RESULT 7		
LOCUS	BG139798	496 bp mRNA linear EST 31-JAN-2001
DEFINITION	EST480240 wild tomato pollen Lycopersicon pennellii cDNA clone	
ACCESSION	CGP14E16 5' sequence, mRNA sequence.	
VERSION	BG139798	
KEYWORDS	BG139798.1 GI:12639986	
SOURCE	EST.	
ORGANISM	Lycopersicon pennellii.	
	Lycopersicon pennellii	
	Eukaryota; Vitidiplantae; Streptophyta; Embryophyta; Tracheophyta;	
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;	
	Asteridae; euasterids I; Solanales; Solanaceae; Solanum;	
	Lycopersicon.	
REFERENCE	1 (bases 1 to 496)	
AUTHORS	van der Hoeven,R., Bezzerides,J., Sun,H., Cho,J., Uteback,T.,	
	Hansen,C., Rönning,C. and Tanksley,S.	
	Generation of ESTs from wild tomato (L. pennellii) pollen	
	unpublished (2001)	
TITLE	Contact: CUGI	
JOURNAL	Clemson University Genomics Institute	
COMMENT	Clemson University	
	100 Jordan Hall, Clemson, SC 29634, USA	
	Email: http://www.genome.clemson.edu/orders/index.html.	

```

source
1. 496
/organism="Lycopersicon pennellii"
/cultivar="TA56"
/db_xref="taxon:28526"
/clone="cLPP14E16"
/clone_lib="wild tomato pollen"
/tissue_type="pollen"
/dev_stage="pollen collected from open flowers"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; Pollen was collected from open flowers from
L.pennellii TA56, and stored at -80 C until library
construction."

BASE COUNT      193 a      93 c      68 g      142 t
ORIGIN

Alignment Scores:
Pred. No.:      1.55      Length:      496
Score:          82.00      Matches:      28
Percent Similarity: 49.37%  Conservative: 11
Best Local Similarity: 35.44%  Mismatches: 27
Query Match:    14.31%  Indels:      13
DB:             12      Gaps:         4

US-09-936-737A-2 (1-103) x BG139798 (1-496)

QY 12 AsnArgLysTyrThr-AspPhe-----AspLysSerPheLysLysSerse 26
    ||| :|||:||||| ||||| ||||| ||||| |||||
Db 262 AATGGCCAAATACACAAAGATTTTATCAAAAATGTAATGATTAAGAAAGATTGGCTTCAAGTGC 321
    ||| :|||:||||| ||||| ||||| ||||| |||||
QY 26 rAspLeuAspGluCysLysThrCysPheLysThrGluTyrCysTyrIleValPheGlu 46
    ||| :|||:||||| ||||| ||||| ||||| |||||
Db 322 ACCAATAGATCATATTAAGAAAGATGTTTAAACA-----TGCAAGAGGTGTATGCA 375
    ||| :|||:||||| ||||| ||||| ||||| |||||
QY 46 uAspThrValasn-----LysGluCysTyrTyrAsnValValAspGluGluLeuAs 64
    ||| :|||:||||| ||||| ||||| ||||| |||||
Db 376 ACATGCATATGATTCATAGAAAAGCAACACAGATGTT-----AAAGACTTAAA 426
    ||| :|||:||||| ||||| ||||| ||||| |||||
QY 64 pGluGluysPheValValAspGluAsnPheThrGluAsnTyrLeuThrAspCys 82
    ||| :|||:||||| ||||| ||||| ||||| |||||
Db 427 TTATTACAAAGCAATATGATATGTCATTAATATACATACCTTGACACTGT 481
    ||| :|||:||||| ||||| ||||| ||||| |||||

RESULT 8
BG136045      529 bp      mRNA      linear      EST 31-JAN-2001
LOCUS      EST476487 wild tomato pollen Lycopersicon pennellii cDNA clone
DEFINITION      cLPP1020 5' sequence, mRNA sequence.
ACCESSION      BG136045
VERSION      BG136045.1 GI:12636233
KEYWORDS      EST.
SOURCE      Lycopersicon pennellii.
ORGANISM      Lycopersicon pennellii.
TITLE      Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
JOURNAL      Eukaryota; Magnoliophyta; eudicotyledons; core eudicots;
COMMENT      Spermatophyta; Asteridae I; Solanales; Solanaceae; Solanum;
      Lycopersicon.
      1 (bases 1 to 529)
      van der Hoeven,R., Bezzerides,J., Sun,H., Cho,J., Uterback,T.,
      Hansen,C., Romning,C. and Tanksley,S.
      Generation of ESTs from wild tomato (L. pennellii) pollen
      Unpublished (2001)
      Contact: CUGI
      Clemson University Genomics Institute
      Clemson University
      100 Jordan Hall, Clemson, SC 29634, USA
      Email: http://www.genome.clemson.edu/orders/index.html.
      Location/Qualifiers
        1..529
        /organism="Lycopersicon pennellii"
        /cultivar="TA56"
        /db_xref="taxon:28526"
        /clone="cLPP1020"
        /clone_lib="wild tomato pollen"

```

```

/tissue_type="pollen"
/dev_stage="pollen collected from open flowers"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; Pollen was collected from open flowers from
L.pennellii TA56, and stored at -80 C until library
construction."

BASE COUNT      196 a      95 c      86 g      152 t
ORIGIN

Alignment Scores:
Pred. No.:      1.67      Length:      529
Score:          82.00      Matches:      26
Percent Similarity: 48.68%  Conservative: 11
Best Local Similarity: 34.21%  Mismatches: 27
Query Match:    14.31%  Indels:      12
DB:             12      Gaps:         4

US-09-936-737A-2 (1-103) x BG136045 (1-529)

QY 14 LysTyrThrAspPhe-----AspLysSerPheLysLysSerseAspLeu 28
    ||| :|||:||||| ||||| ||||| ||||| |||||
Db 238 AAGACACAAAGATTTTATCAAAAATGTAATGATTAAGAAAGATTGGCTTCAAGTCAACCAATA 297
    ||| :|||:||||| ||||| ||||| ||||| |||||
QY 29 AspGluCysLysLysThrCysPheLysThrGluTyrCysTyrIleValPheGluAspThr 48
    ||| :|||:||||| ||||| ||||| ||||| |||||
Db 298 GATCATTTATAGAAAGATGTTTAAACA-----TGCAAGAGGTGTATGCAAGATGCA 351
    ||| :|||:||||| ||||| ||||| ||||| |||||
QY 49 Valasn-----LysGluCysTyrTyrAsnValValAspGluGluLeuAspGluGlu 66
    ||| :|||:||||| ||||| ||||| ||||| |||||
Db 352 ATTGATTCATAGAAAAGCAACACAGATGTT-----AAAGACTTAAATTTATAC 402
    ||| :|||:||||| ||||| ||||| ||||| |||||
QY 67 LysPheValValAspGluAsnPheThrGluAsnTyrLeuThrAspCys 82
    ||| :|||:||||| ||||| ||||| ||||| |||||
Db 403 AAAGCAATATGATATGTCATCATATACATACCTTGACACTGT 450
    ||| :|||:||||| ||||| ||||| ||||| |||||

RESULT 9
AW737851      541 bp      mRNA      linear      EST 18-MAY-2001
LOCUS      EST339278 tomato flower buds, anthesis, Cornell University
DEFINITION      Lycopersicon esculentum cDNA clone CTOD4B1 5', mRNA sequence.
ACCESSION      AW737851
VERSION      AW737851.1 GI:7646796
KEYWORDS      EST.
SOURCE      tomato.
ORGANISM      Lycopersicon esculentum
TITLE      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
JOURNAL      Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
COMMENT      Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
      Lycopersicon.
      1 (bases 1 to 541)
      van der Hoeven,R.S., Bezzerides,J.L., Matern,A.L., Holt,I.E., Liang
      F., Hansen,T.S., Craven,M.B., Bowman,C.L., Romning,C.M., Nierman
      W., Fraser,C.M., Martin,G.B., Giovannoni,J.J., and Tanksley,S.D.
      Generation of ESTs from tomato flower tissue, anthesis
      Unpublished (1999)
      Contact: CUGI
      Clemson University Genomics Institute
      Clemson University
      100 Jordan Hall, Clemson, SC 29634, USA
      Email: http://www.genome.clemson.edu/orders/index.html
      5 prime sequence.
      Location/Qualifiers
        1..541
        /organism="Lycopersicon esculentum"
        /cultivar="TA496"
        /db_xref="taxon:4081"
        /clone="CTOD4B1"
        /clone_lib="tomato flower buds, anthesis, Cornell
        University"
        /tissue_type="flower"
        /dev_stage="anthesis"
        /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:

```

Xhol: supplier: Tanksley; Flower buds and flowers were taken from greenhouse plants (4-8 wks old, TA96). They were immediately frozen in liquid nitrogen and then size-separated while remaining frozen."

BASE COUNT

196 a 96 c 90 g 159 t

Alignment Scores:

Pred. No.:	1.72	Length:	541
Score:	82.00	Matches:	26
Percent Similarity:	48.68%	Conservative:	11
Best Local Similarity:	34.21%	Mismatches:	27
Query Match:	14.31%	Indels:	12
DB:	10	Gaps:	4

US-09-936-737A-2 (1-103) x AM737851 (1-541)

OY 14 LysTyrThrAspPhe-----AspLysSerPheLysSerSerAspLeu 28
 DB 225 AAGACACAGATTTTATCAAGAAATGTAATGATATAAGATTGGCTTCAAGTCACCAATA 284
 OY 29 AspLysCysLysLysThrCysPheLysThrGluTyrCysTyrIleValPheGluAspThr 48
 DB 285 GATCATTAACAAAAAGATGTTTAAACA-----TGCAGAGAGGTGATGAGATGCA 338
 OY 49 ValAsn-----LysGluCysTyrTyrAsnValValAspGlyGluGluLeuAspGlnGlu 66
 DB 339 ATTGATTCATGAAAGAACACACAGATGTT-----AAGAACTAATATATTAC 389
 OY 67 LysPheValValAspLysLysPheThrGluAsnThrGluAsnTyrLeuThrAspCys 82
 DB 390 AAGCAAAATATGATATGTTGTCATCATATACATACCTTGACACTTGT 437

RESULT 10

LOCUS BG139905 548 bp mRNA linear EST 31-JAN-2001
 DEFINITION EST480347 wild tomato pollen Lycopersicon pennellii cDNA clone
 ACCESSION BG139905
 VERSION BG139905.1 GI:12640093
 KEYWORDS EST.
 SOURCE Lycopersicon pennellii.
 ORGANISM Lycopersicon pennellii.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.
 1 (bases 1 to 548)
 van der Hoeven, R., Bezzerides, J., Sun, H., Cho, J., Utterback, T., Hansen, C., Roming, C. and Tanksley, S.
 Generation of ESTs from wild tomato (L. pennellii) pollen
 Unpublished (2001)
 Contact: CUGI
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Email: <http://www.genome.clemson.edu/orders/index.html>.
 Location/Qualifiers
 1..548

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

FEATURES
 source

/organism="Lycopersicon pennellii"
 /cultivar="TA56"
 /db_xref="taxon:28526"
 /clone="CLP14N18"
 /clone_lib="wild tomato pollen"
 /tissue_type="pollen"
 /dev_stage="pollen collected from open flowers"
 /lab_host="SOLR"
 /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; Pollen was collected from open flowers from L. pennellii TA56, and stored at -80 C until library construction."
 BASE COUNT
 ORIGIN
 199 a 99 c 87 g 163 t

Alignment Scores:

Pred. No.:	1.75	Length:	548
Score:	82.00	Matches:	26
Percent Similarity:	48.68%	Conservative:	11
Best Local Similarity:	34.21%	Mismatches:	27
Query Match:	14.31%	Indels:	12
DB:	12	Gaps:	4

US-09-936-737A-2 (1-103) x BG139905 (1-548)

OY 14 LysTyrThrAspPhe-----AspLysSerPheLysSerSerAspLeu 28
 DB 225 AAGACACAGATTTTATCAAAATGTAATGATATAAGATTGGCTTCAAGTCACCAATA 284
 OY 29 AspLysCysLysLysThrCysPheLysThrGluTyrCysTyrIleValPheGluAspThr 48
 DB 285 GATCATTAACAAAAAGATGTTTAAACA-----TGCAGAGAGGTGATGAGATGCA 338
 OY 49 ValAsn-----LysGluCysTyrTyrAsnValValAspGlyGluGluLeuAspGlnGlu 66
 DB 339 ATTGATTCATGAAAGAACACACAGATGTT-----AAGAACTAATATATTAC 389
 OY 67 LysPheValValAspLysLysPheThrGluAsnThrGluAsnTyrLeuThrAspCys 82
 DB 390 AAGCAAAATATGATATGTTGTCATCATATACATACCTTGACACTTGT 437

RESULT 11

LOCUS BG136094 561 bp mRNA linear EST 31-JAN-2001
 DEFINITION EST476536 wild tomato pollen Lycopersicon pennellii cDNA clone
 ACCESSION BG136094
 VERSION BG136094.1 GI:12636282
 KEYWORDS EST.
 SOURCE Lycopersicon pennellii.
 ORGANISM Lycopersicon pennellii.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.
 1 (bases 1 to 561)
 van der Hoeven, R., Bezzerides, J., Sun, H., Cho, J., Utterback, T., Hansen, C., Roming, C. and Tanksley, S.
 Generation of ESTs from wild tomato (L. pennellii) pollen
 Unpublished (2001)
 Contact: CUGI
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Email: <http://www.genome.clemson.edu/orders/index.html>.
 Location/Qualifiers
 1..561

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

FEATURES
 source

/organism="Lycopersicon pennellii"
 /cultivar="TA56"
 /db_xref="taxon:28526"
 /clone="CLP1J21"
 /clone_lib="wild tomato pollen"
 /tissue_type="pollen"
 /dev_stage="pollen collected from open flowers"
 /lab_host="SOLR"
 /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; Pollen was collected from open flowers from L. pennellii TA56, and stored at -80 C until library construction."
 BASE COUNT
 ORIGIN
 208 a 99 c 90 g 164 t

Alignment Scores:
 Pred. No.: 1.8
 Score: 82.00
 Percent Similarity: 48.68%
 Best Local Similarity: 34.21%
 Length: 561
 Matches: 26
 Conservative: 11
 Mismatches: 27

Query Match:	14.31%	Indels:	12
DB:	12	Gaps:	4
US-09-936-737A-2 (1-103) x BG136094 (1-561)			
Qy 14	LysTyrThrAspPhe-----	AspLysSerPheLysLysSerSerAspLeu 28	
Db 238	NAGACACAAAGTTTATTCACAAAATGCTATTCATATAAAGATGGCTTCAGAGCACCAATTA 297		
Qy 29	AspGluCysLysLysThrCysPheLysThrGluTyrCysTyrLleValPheGluAspThr 48		
Db 298	GATCATATTAAAGAAAGATGTTGTTTAAACA-----TCCAAAGAGGTATGAAATCA 351		
Qy 49	ValAsn-----LysGluCysTyrTyrAsnValValAspGlyGluGluLeuAspGlnGlu 66		
Db 352	ATTGATTCAATGAAAAAGCAACACACAGATGTT-----AAAGAACTAAATTATTAC 402		
Qy 67	LysPheValValAspGluAsnPheThrGluAsnTyrLeuThrAspCys 82		
Db 403	AAAGCAATATTCGATTATTCGTCATCATACATACATACCTTGACACCTTGT 450		
RESULT 12			
LOCUS	BG138707	561 bp	linear
DEFINITION	EST479149 wild tomato pollen Lycopersicon pennellii cDNA clone		EST 31-JAN-2001
ACCESSION	CP101P12	5' sequence, mRNA sequence.	
VERSION	BG138707		
KEYWORDS	BG138707.1 GI:12638895		
SOURCE	EST.		
ORGANISM	Lycopersicon pennellii.		
REFERENCE	Lycopersicon pennellii.		
AUTHORS	Eukaryota; Vitidiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.		
TITLE	1 (bases 1 to 561)		
JOURNAL	van der Hoeven, R., Bezzerides, J., Sun, H., Cho, J., Utterback, T., Hansen, C., Rönning, C. and Tanksley, S.		
COMMENT	Generation of ESTs from wild tomato (L. pennellii) pollen Unpublished (2001)		
FEATURES	Contact: CUGI		
source	Clemson University Genomics Institute Clemson University 100 Jordan Hall, Clemson, SC 29634, USA Email: http://www.genome.clemson.edu/orders/index.html . Location/Qualifiers 1. 561 /organism="Lycopersicon pennellii" /cultivar="TA56" /db_xref="taxon:28526" /clone="CLP10P12" /clone_lib="wild tomato pollen" /tissue_type="pollen" /dev_stage="pollen collected from open flowers" /lab_host="SOLR" /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI. Pollen was collected from open flowers from L. pennellii TA56, and stored at -80 C until library construction."		
BASE COUNT	208 a 99 c 90 g 164 t		
ORIGIN			
Alignment Scores:			
Pred. NO.:	1.8	Length:	561
Score:	82.00	Matches:	26
Percent Similarity:	48.68%	Conservative:	11
Best Local Similarity:	34.21%	Mismatches:	27
Query Match:	14.31%	Indels:	12
DB:	12	Gaps:	4
US-09-936-737A-2 (1-103) x BG138707 (1-561)			
Qy 14	LysTyrThrAspPhe-----	AspLysSerPheLysLysSerSerAspLeu 28	

Db	238	AAGACACAAGATTATTCATAAAANGTAATTCATATAAAGATTGGCTCAAGTACCACCAATA	297
Oy	29	ASPGlucylAspGlyThrCysPheIysThrgIuTrcYstfrrileValPheGIuaSpthr	48
Db	298	GATCATTTATACGAAGAAGATTGTTTAAAACA-----TGCAAAGAGGCGTATGAAGATGCA	351
Oy	49	ValAsn-----LysGIuCysfYrTrasnValAlaSPglugluLeuaspGinglu	66
Db	352	ATTGATTCAATGAAAAAGCAACACAGATGTT-----AAAGACTAAATTTATAC	402
Oy	67	LysPheValValaSPgluAsnPherThrgIuAsnTrIeuThrAspCys	82
Db	403	AAGCAAAATATGGATTTGGTCGACATATACATACCTTGACACTGT	450
RESULT 13	BGI39936	561 bp	linear EST 31-JAN-2001
LOCUS	EST480378	wild tomato pollen Lycopersicon pennellii	cDNA clone
DEFINITION	clpP15g 5' sequence, mRNA sequence.		
ACCESSION	BGI39936		
VERSION	BGI39936.1	GI:12640124	
SOURCE	EST.		
ORGANISM	Lycopersicon pennellii.		
REFERENCE	Lycopersicon pennellii.		
AUTHORS	Eukaryota; Vitidiplantae; streptophyta; Embryophyta; Tracheophyta;		
TITLE	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;		
COMMENT	Asteridae; euasterids I; Solanales; Solanaceae; Solanum;		
FEATURES	Lycopersicon.		
source	1. (bases 1 to 561)		
	van der Hoeven,R., Bezzerides,J., Sun,H., Cho,J., Utterback,T.,		
	Hansen,C., Romling,C. and Tanksley,S.		
	Generation of ESTs from wild tomato (L. pennellii) pollen		
	Unpublished (2001)		
	Contact: CUGI		
	Clemson University Genomics Institute		
	Clemson University		
	100 Jordan Hall, Clemson, SC 29634, USA		
	Email: http://www.genome.clemson.edu/orders/index.html.		
	Location/Qualifiers		
	/organism="Lycopersicon pennellii"		
	/cultivar="TA56"		
	/db_xref="taxon:28526"		
	/clone="clpP15g3"		
	/clone_lib="wild tomato pollen"		
	/tissue_type="pollen"		
	/dev_stage="pollen collected from open flowers"		
	/lab_host="SOLR"		
	/note=vector: pBluescript SK(+); site_1: EcoRI; Site_2:		
	XhoI; Pollen was collected from open flowers from		
	L. pennellii TA56, and stored at -80 C until library		
	construction."		
BASE COUNT	208 a	99 c	165 t
ORIGIN			
Alignment Scores:			
Pred. No.:	1.8	length:	561
Score:	82.00	Matches:	26
Percent Similarity:	48.68%	Conservative:	11
Best Local Similarity:	34.21%	Mismatches:	27
Query Match:	14.31%	Indels:	12
DB:	12	Gaps:	4
US-09-936-737A-2 (1-103) x BGI39936 (1-561)			
Oy	14	LysTyrtThrAsphe-----AspYssSerPheIysSerSerAspIeu	28
Db	238	AAGACACAAGATTATTCATAAAATGTATATGAATAAAGATTGGCTCAAGTACCACCAATA	297
Oy	29	ASPGlucylAspGlyThrCysPheIysThrgIuTrcYstfrrileValPheGIuaSpthr	48
Db	298	GATCATTTATACGAAGAAGATTGTTTAAAACA-----TGCAAAGAGGCGTATGAAGATGCA	351

